INVESTIGATING GENETIC SUSCEPTIBILITY TO *RHODOCOCCUS EQUI*

A Dissertation

by

COLE MATTHEW MCQUEEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Noah Cohen
Co-Chair of Committee,	Scott Dindot
Committee Members,	Allen Roussel
	Jason Sawyer
Intercollegiate Faculty Chair,	Craig Coates

December 2015

Major Subject: Genetics

ABSTRACT

Rhodococcus equi causes pneumonia and extrapulmonary disorders in foals and other immunocompromised animals including people. Although exposure to virulent *R. equi* is widespread in the environment of foals, only a small proportion of foals develop *R. equi* pneumonia at affected farms. It remains unclear why some foals develop disease while other exposed foals do not. Anecdotal evidence suggests there may be an underlying genetic predisposition to disease resulting from *R. equi* infection and the genetic contributions of the host remain ill-defined.

A genome wide association study (GWAS) examining single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) to identify regions of the genome contributing to disease was performed. Although no associations were made with CNVs, a set (*viz.,* four) of SNPs were significantly associated with *R. equi* pneumonia. The SNPs identified localized to a region on chromosome 26 containing a potential candidate gene. The transient receptor potential cation channel, subfamily M, group 2 (TRPM2) gene is known to play a critical role in neutrophil recruitment and the severity of tissue damage at sites of inflammation. One associated SNPs in the GWAS is located within this gene and was used, via PCR, to validate the GWAS finding in a joint analysis. Joint analysis revealed a 3- to 4-fold increase in odds of disease for individuals homozygous for the SNP identified in TRPM2. This finding was consistent across each clinical group compared.

RNA-Seq was performed to further investigate this region, as well as identify others across the genome, and understand the functional implications on gene expression that may be marked by the identified SNPs. Each *TRPM2* genotype was represented (*viz.*, AA, AB, BB) in the 12 horses that were donors for sequencing. RNA-Seq analyses identified several novel transcripts across the *TRPM2* region; however, none were differentially expressed in relation to the *TRPM2* genotypes. Several genes identified as being differentially expressed were linked through pathway analysis which further implicated innate immunity as being a critical player in the pathogenesis of *R. equi* pneumonia. Further studies are required to identify targets and practices for the control and prevention of *R. equi* pneumonia in foals.

DEDICATION

I would like to thank my wife, children, and parents for their continuous support and encouragement throughout my pursuit of a doctoral degree.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Cohen, and my committee members, Dr. Dindot, Dr. Roussel, and Dr. Sawyer, for their guidance and support throughout the course of this research.

I also want to extend my gratitude to the following: the United States Department of Agriculture, the LINK Equine Endowment, the Department of Large Animal Clinical Sciences, the College of Veterinary Medicine, the Interdisciplinary Department of Genetics, the Morris Animal Foundation, and the 6666 Ranch. All the above listed entities provided in some form: funding, travel, samples, and educational support for my education. Also, I would like to thank the Texas A&M Institute for Genome Sciences and Society for computing and technical support.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION: GENETIC SUSCEPTIBILITY TO *RHODOCOCCUS EQUI**

Introduction

Rhodococcus equi pneumonia is an important disease of foals most commonly characterized by chronic progression associated with development of large pulmonary abscesses [1]. Treatment of *R. equi* pneumonia is prolonged and expensive, and prevention is limited because transfusion of hyperimmune plasma is incompletely effective [2], chemoprophylaxis is inconsistently effective [3,4] and may promote antimicrobial resistance [5], and no effective vaccine is currently available [6]. Isolates of *R. equi* that are virulent in foals express the virulence-associated protein A (VapA), which is encoded by a gene, located within a pathogenicity island, on an approximately 85- to 90-kilobase (kb) plasmid. Expression of VapA alone however is not sufficient to cause disease [7,8]. Many different strains of virulent (and avirulent) *R. equi* have been shown to be present in a common environment (*i.e.,* the same farm), and multiple genotypic virulent strains may exist even in an individual foal with *R. equi* pneumonia [5,9-11]. Although exposure to *R. equi* is widespread at farms where foals are affected, only a variable proportion of foals will develop clinical disease at a given farm whereas other foals at the same location will not develop disease [12,13]. In addition, anecdotal reports by

*Reprinted with permission from "Genetic Susceptibility to *Rhodococcus equi*" by McQueen et al., 2015. *JVIM*, doi: 10.1111/jvim.13616, 2015 by CM McQueen.

veterinarians indicate that some mares recurrently have affected foals whereas other mares from the same environment consistently have foals that do not develop *R. equi* pneumonia. Taken together, these findings support the possibility of an important role for a genetic predisposition (*i.e.*, susceptibility, resistance, or tolerance) to development of *R. equi* pneumonia and have prompted investigations of the genetic basis for this disease. Pneumonia caused by *R. equi* is a complex trait. Thus, it is unlikely that it will have a monogenic basis. Nevertheless, studying the genetic basis of *R. equi* pneumonia is important because it could reveal information about crucial biological processes and pathways that influence the outcome of infection in foals, and identifying these pathways and processes might consequently lead to novel approaches for diagnostic testing, treatment, control and prevention. The purpose of this report is to review what is known about genetic predisposition to pneumonia caused by *R. equi* in foals and describe genetic techniques currently available to study the genetic determinants of development of *R. equi* pneumonia or other diseases in horses. We begin by summarizing the current literature regarding genetic associations with *R. equi* pneumonia, and then discuss some more advanced genetic tools available for future studies to further investigate the genetic basis of *R. equi* pneumonia.

Literature search

A literature search was conducted to identify English language studies from any year that focused on foals, *R. equi*, and genetics. Databases were searched in April 2014 through Ovid including CAB Abstracts, MEDLINE, Embase, and BIOSIS. Search words

included (foals or equus or equine) and (r equi or *Rhodococcus equi**) and gene*, where the asterisk indicates truncation. Within each database, appropriate subject headings or index terms also were added. A total of 744 articles were retrieved and de-duplicated, with 5 articles selected for inclusion. This search was updated in September 2014.

Candidate genes

We identified 5 studies that attempted to identify genes associated with *R. equi* pneumonia. Four of these 5 studies have utilized a candidate gene approach (Table 1.1). The candidate gene approach involves either use of prior knowledge pertaining to known gene functions that might predispose to the disease of interest (*e.g.,* the interferon-gamma pathway and *R. equi* pneumonia) [14,15] or use of genes implicated in other but similar diseases that could be potential candidates for involvement(*e.g.,* genes important for *Mycobacterium tuberculosis* which, like *R. equi,* is a gram-positive, facultative intracellular organism that replicates primarily within macrophages and causes pneumonia and could be potential candidates for *R. equi* pneumonia) [16-18]. To the authors' knowledge, the first candidate gene association study of *R. equi* pneumonia compared the frequencies of single nucleotide polymorphisms (SNPs) in the transferrin gene (*Tf)* among Thoroughbred foals from Kentucky that died of *R. equi* pneumonia with those of control Thoroughbred mares [19]. In one study, the *Tf* gene was selected on the basis of its product's ability to bind iron because iron sequestration is a known

Table 1.1. Genetic studies of *Rhodococcus equi* pneumonia.

Author	Study Design	Country	Breed(s)	Number of Horses	Markers Investigated	Observed Outcome	Findings
Mousel *et al.,*	Candidate gene	USA	Thoroughbred	N = 84	*Tf* SNPs	Clinical pneumonia or control	Allelic association of *Tf* with disease
Horin *et al.,*	Candidate gene	Czech Republic	Thoroughbred	N = 51	SNPs, Microsatellites	Burden of *R. equi* in TBA fluid	Association of *IL1RN and IL1β* with burden of *R. equi*
Halbert *et al.,*	Candidate gene	USA	Arabian and Thoroughbred	N = 103	SLC11A1 SNPs	Clinical pneumonia or control	Variation in *SLC11A1* associated with disease
Horin *et al.,*	Candidate gene	Czech Republic	Thoroughbred	N = 51	SNPs	Burden of *R. equi* in TBA fluid	Association of *IL7R* with burden of *R. equi*
McQueen *et al.,*	GWAS	USA	Quarter Horse	N = 72 N=248	Genome wide SNPs	Clinical and Subclinical pneumonia, or control	Associated SNP in *TRPM2* with disease

host defense mechanism against bacterial replication [20,21]. The authors postulated that polymorphisms in the *Tf* gene might result in enhanced (or decreased) iron binding, which then could confer a selective advantage (or disadvantage) to survive infections with bacteria such as *R. equi* [19]. The authors used SNP frequencies to infer *Tf* alleles present within the study population, and allele frequencies were subsequently compared between the case and control groups. The authors documented a significant ($P < 0.05$) abundance of the *Tf F* allele and a deficiency of the D_1 allele among the cases (diseased foals) when compared with controls. Limitations of this study included the fac that sample size was relatively small, it was restricted to a single breed, a separate population for validation was not included, and no mechanistic studies (*i.e.,* documentation that the *F* allele was associated with decreased iron-binding) were incorporated or cited. Nonetheless, a significant association of SNPs in the *Tf* gene with *R. equi* pneumonia was demonstrated, and this finding represented an important advance in knowledge.

A later study seeking to identify a genetic predisposition to *R. equi* pneumonia utilized the candidate gene approach by comparing frequencies of 22 genetic markers among 51 Thoroughbred foals from the Czech Republic [22]. These markers were either SNPs or polymorphic microsatellites in or near immune-related genes that had been previously identified (except for 5 markers that were first identified in this study). No genetic variants were significantly associated with *R. equi* pneumonia, but some genetic variants were significantly associated with a higher burden of *R. equi* in tracheobronchial aspirate

(TBA) fluid from foals. Specifically, loci on chromosome 10 and 15 were associated with *R. equi* infection when comparing the subset of foals with extreme phenotypes (*i.e.*, foals with the highest numbers of *R. equi* in TBA fluid) to those with no *R. equi*. The strongest association with TBA fluid phenotype was for the microsatellite HMS01 located on chromosome 15 which encodes the genes for interleukin (IL)-1β (*IL1β*) and the IL-1 receptor antagonist (*IL1RN*). Although the associations were relatively weak and the phenotype was for burden of *R. equi* in TBA fluid (rather than for pneumonia caused by *R. equi*), these results offer further evidence of a genetic basis for host response to infection with *R. equi*.

A third study from our laboratory utilized previous findings indicating association between the solute carrier family 11 member 1 gene (*SLC11A1*) and susceptibility to intracellular bacterial infections in other species of animals [23-26]. The *SLC11A1* gene encodes a protein relevant to innate immune responses to intracellular bacteria [27,28]. Direct sequencing of the beginning of the gene transcript (*i.e.*, the 5′ end of the gene) was used to identify SNPs that were compared between cases of *R. equi* and unaffected foals (controls) among Arabian horses at 2 farms (1 farm in Texas and 1 farm in Arizona). A novel SNP, -57T, in the 5′ untranslated region (UTR) was significantly associated with *R. equi* pneumonia [29]. The authors further screened for this polymorphism in 5 domestic horse breeds, donkeys, and zebras, and found it was represented in 4 of the 5 horse breeds. The observation that this SNP was represented across multiple breeds strengthened the study's findings because if a marker were

present in only 1 breed, it would be possible (if not probable) that the identified marker was more likely associated with breed differences than disease. Limitations of this study included the fact that association between the candidate gene and disease was only assessed within a single breed at 2 farms, and no validation of the association in another population was conducted. Additionally, inconsistencies in diagnostic practices for *R. equi* pneumonia among farms in the study existed, which might have impacted the results [29].

A fourth candidate gene study investigated the frequency of SNPs in selected immune response genes from DNA samples collected from 31 Thoroughbred foals from the Czech Republic [30] that had been used in a previous candidate gene study (described above) [22]. The candidate markers were used to asses allele frequencies between groups of foals classified on the basis of a binary outcome using a cut-point of > 5,000 colony forming units (CFU)/mL of *vapA*-positive *R. equi* in TBA fluid. Twenty-five foals were categorized as below the cut-point because they had no bacteria cultured from them, and 6 were categorized as above the cut-point [30]. An association was identified between a SNP in the IL-7 receptor (*IL7R*) gene and the presence of > 5,000 CFU of *R. equi* in TBA fluid. Limitations of this study included lack of a validation population in which the association could be replicated, and, similar to the earlier study using these foals, the association was not made between the marker and disease but rather between the marker and the concentrations of bacteria present in TBA fluid. Regardless of these limitations, this study was scientifically important in that it implicated the *IL7R* gene in particular,

and innate immunity in general, as having a role in host response to infection with *R. equi*.

The candidate gene approach is a valid method for genetic investigation and yielded positive associations in the aforementioned studies, strengthening the plausibility of a genetic contribution to susceptibility or resistance to *R. equi* pneumonia in foals. Moreover, the commonality of identifying innate immune responses as playing a role in host defense against infection with *R. equi* in these various candidate gene association studies is important to our understanding of the pathogenesis of *R. equi* pneumonia.

Despite these positive results, the candidate gene approach has important limitations for making genetic associations. Bias is introduced into the study design by selecting a small number of genes for evaluation, on the basis of either function of the gene product(s) or prior association of the gene with disease. This selection process effectively eliminates the ability of the investigators to examine both the enormous amount of genetic information in the remainder of the genome or the relationship and interaction of other genes with the genes of interest [17]. Other genetic elements present in the genome (*e.g.,* sites critical to gene regulation) are missed by restricting analysis to candidate genes, because in most cases probes used to detect variation are not near each other and offer no information about neighboring genetic variation. Assessing variation across the genome circumvents these limitations of the candidate gene approach.

Genome-wide studies in horses are now feasible because of recent technological developments.

Genome-wide association studies

Genome-wide association studies (GWASs) rapidly gained popularity after the sequencing of the genomes of several animal species, including human beings [31-34]. The completion of the sequencing and assembly of these reference genomes (an assembly of the DNA sequence and its chromosomal locations representing the genetic baseline of a species) provided a tool that could be used as a map indicating where elements of the genome reside. Substantial efforts then were made to catalogue the locations of genes and genetic variation identified within these species [35,36]. Single nucleotide polymorphisms proved useful for characterizing genetic variation among individuals of a given species, and the development of SNP array technology made it possible to perform > 1 million association tests simultaneously of markers across the entire genome without the expense or labor of genome sequencing.

Single nucleotide polymorphism arrays are glass slides with genomic probes (sequences of DNA) that capture SNPs present within a given species. Through a hybridization process, the probes bind DNA of samples to identify which polymorphisms are present in that sample [37]. These SNP arrays enabled clinical researchers to compare clinically affected horses with unaffected controls so as to examine the association of various health disorders with markers on a genome-wide basis, and the interplay among different

genetic variants associated with disease [38]. Results from a GWAS are readily identifiable because they typically are visualized by plotting the negative logarithm of the P value for the association of a given SNP with the outcome of interest as the ordinate (vertical axis) and the chromosome number as the abscissa (horizontal axis). The resultant scatter plots are known as Manhattan plots because they resemble the skyline of a major city with some points that tower over the majority of others. Determining the genome sequence of the domestic horse led to the development of 2 equine SNP arrays that could be used for GWAS by researchers [32,39].

Currently, a single SNP array has been developed, well characterized, made commercially available, and utilized in numerous GWAS in horses. For example, the EquineSNP70 BeadChip Array[a] contains approximately 74,000 SNPs positioned across the equine genome that can be simultaneously tested to identify their associations with a phenotype of interest [39,40]. Recently, a higher density SNP array with approximately 770,000 SNPs across the equine genome has been developed but has not been characterized in peer-reviewed literature to date. Several GWAS in horses using SNP arrays and yielding positive associations have been reported [41-46]. Genome-wide association studies rely on observing different frequencies of alleles (identified by SNPs) that segregate with a phenotype of interest. These associations have identified regions of interest (Figure 1.1 A), which are further investigated to understand what elements

[a] EquineSNP70 BeadChip ArrayIllumina, San Diego, California, USA

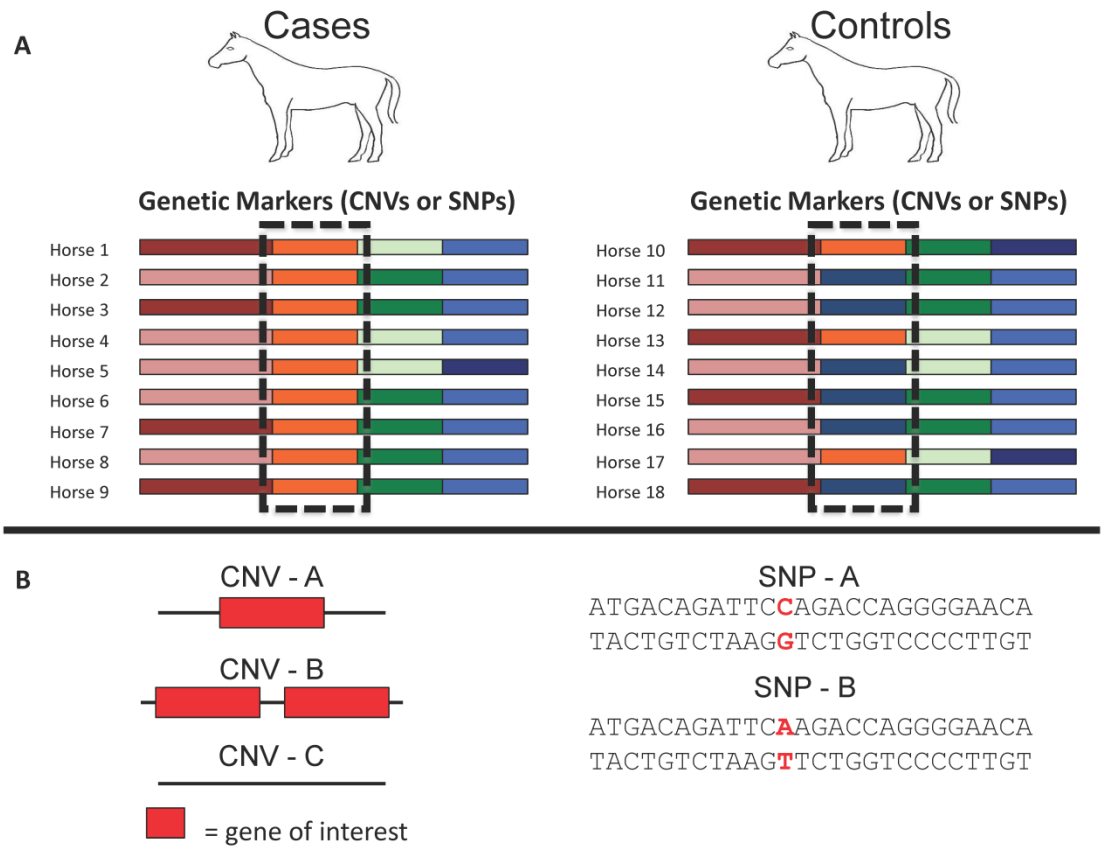


Figure 1.1. Association studies, CNVs, and SNPs. (**A**) The colored blocks indicate different alleles or haplotypes present in the horse genome. These have been identified by either a CNV or a SNP but any type of genetic variation can be used to identify alleles. The boxed regions show a greater frequency of the orange allele in the cases compared to the controls. The increased frequency of this allele in the cases suggests that it may harbor a variant(s) causing or contributing to the associated phenotype. (**B**) CNV – A represents a single copy of a gene; CNV – B represents a duplication of the gene; and, CNV – C represents a deletion of the gene. These examples demonstrate how CNVs can affect a single gene and can be used to identify different alleles in a population. SNPs, represented as red bases, offer the ability to identify alleles because of their polymorphic nature. Either type of genetic variation can be used in a GWAS to identify disease associated alleles.

(*e.g.,* genes, promoters, other variants), pathways or processes are associated with the phenotype.

The reason marker associations require region investigation is because of linkage disequilibrium (LD), which is defined as the non-random association of genetic information [47]. Linkage disequilibrium is a phenomenon that allows for the prediction of the non-genotyped genetic information around a genotyped marker because of an assumption that the genetic material around a marker differs and thus can be based on the allele represented by the marker (*i.e.,* SNP). The use of LD to make disease associations leverages inheritance patterns, selection, and evolution and is a fundamental concept underlying GWAS. The association of a marker, whether it is located in a gene or in a noncoding region of the genome, should only be treated as an indicator of the need to further investigate the area. An association of a SNP with disease neither indicates that the SNP is causally associated with the disease nor that the specific gene in which the SNP lies is the gene of interest. A SNP only indicates that there might be genetic variation in the area of the genome where the SNP is located. The size of the area of interest is largely described by the length of the LD (i.e., the number of bases for which another gene or genetic element can be expected to be in LD with the marker). Using LD to assist in making associations is a powerful tool that is genome-wide and efficient because not all markers across a genome must be tested to find an associated region, should one exist. The power of LD allows fewer markers to be present on an array, and hence decreases the number of necessary test corrections. Furthermore, the

longer the LD of the species, the fewer SNPs are necessary to identify significant associations. Estimates of LD for breeds of horses are markedly longer than those for humans [48,49]. Thus, one might expect to need fewer SNPs on an equine array to have the same discriminatory power as a human array or to have greater power in a GWAS for horses than humans for an array of the same size or density of SNPs.

Although SNP arrays are proving to be a powerful tool for investigating the relationship of genetic and phenotypic variation, challenges exist with validating and reproducing results generated by SNP-based GWAS. There are likely many contributing risk alleles for all complex traits and complex diseases such that no single allele can explain all of the phenotypic variation [50]. This becomes problematic during replication using different breeds and populations because the markers identified might merely reflect breed differences, or the markers might represent different alleles conferring different levels of risk across breeds or populations. The number of association studies in equine genetics will only continue to increase and the equine research community will continue to face these challenges. Appropriate study designs, accurately defined and categorized phenotypes, and functional follow-up assays will be essential to maximize the utility of GWAS results in future studies [51].

The first report of a GWAS with *R. equi* pneumonia recently was published by our laboratory [52]. The study [53] population included 248 foals born in 2011 at a large Quarter Horse breeding farm. For a separate study characterizing the accuracy of screening tests

for *R. equi* pneumonia, foals at the farm were examined by thoracic ultrasonography every 2 weeks beginning at 3 weeks of age and continuing through 19 weeks of age (or until weaned) to identify foals with areas of pulmonary consolidation or abscess formation attributed to *R. equi* infection. Farm personnel were blinded to the ultrasonographic findings and a separate team of individuals performed thoracic ultrasonography. Foals at the farm were classified as having *R. equi* pneumonia (N = 43; on the basis of clinical signs of pneumonia, isolation of virulent *R. equi* from the TBA fluid, cytologic evidence of sepsis in TBA fluid, and ultrasonographic evidence of pulmonary consolidation or abscess formation > 1 cm in maximal diameter), no pneumonia (N = 49; on the basis of absence of clinical signs of pneumonia and no ultrasonographic evidence of pulmonary consolidation or abscess formation > 1 cm diameter), and subclinical pneumonia (N = 156; on the basis of absence of clinical signs of pneumonia with ultrasonographic evidence of pulmonary consolidation or abscess formation > 1 cm diameter). From each of these 3 subpopulations of foals, a sample of 24 foals was randomly selected for genotyping using the EquineSNP70 BeadChip Array. Comparisons among the 3 groups identified a significant association of a region on chromosome 26 that included the gene for the transient receptor potential cation channel, subfamily M, member 2 (*TRPM2*).

These results are notable because the *TRPM2* gene is known to play a role in neutrophil function and recruitment. In a study using *TRPM2* knock-out mice and a model of ulcerative colitis, *TRPM2*-deficient mice had less tissue damage at sites of inflammation

than did wild-type mice [54]. The association of the *TRPM2* was validated using polymerase chain reaction (PCR)-based genotyping of the locus in the remaining 176 study foals that were not tested using the SNP array. The principal limitations of this study were that only a single breed at a single farm was studied, and that no association of the genotype with function of the *TRPM2* gene-product or associated signaling pathways was identified. Nonetheless, this study is interesting in that, consistent with previous candidate gene studies, a gene related to innate immunity was associated with *R. equi* pneumonia, and the study provides further evidence of the underlying genetic basis for *R. equi* pneumonia.

Copy number variants

Although the genetic determinants of phenotypic variation are largely dependent on the gene or genes and the manner in which they exert their effect (e.g., altering the biochemical properties of a protein, changing the expression patterns or levels of messenger RNA), recent studies have implicated copy number variants (CNVs) as major determinants of phenotypic variation in humans and animals [55-57]. As the name implies, CNVs are characterized by changes in the number of copies of DNA between at least 2 individuals (Figure 1.1 B) [58]. Their sizes can range from hundreds to millions of base-pairs (bps). Although they often are enriched in certain regions of the genome that predispose to their formation, CNVs have been detected throughout the genome, with many CNVs involving multiple genes, individual genes, or components of a single gene.

Several mechanisms have been shown to cause the formation of CNVs. During meiosis in the germ cells, homologous chromosomes align with each other to exchange genetic information between the parental genomes. This process, called homologous recombination or crossing over, plays an instrumental role in expanding the genetic diversity of a population. In rare instances however, the exchange of genetic material can occur between 2 different sites (non-allelic homologous recombination [NAHR]), resulting in an unequal exchange of genetic material [59]. Although NAHR often is the source of many disease-causing CNVs, this process plays a key role in the formation of gene families and the birth of new genes. Naturally-occurring DNA repair mechanisms also can delete or duplicate DNA. For example, the non-homologous end joining (NHEJ) and microhomology mediated end-joining (MMEJ) pathways are used to repair double-stranded DNA breaks that occur in the genome. During the repair of the breaks, the NHEJ and MMEJ pathways either add or remove DNA to ligate the broken strands back together [60-62]. Fork stalling and template switching (FoSTeS) is a mechanism used to circumvent a stalled replication complex during DNA synthesis. When this happens, the FoSTeS machinery identifies a similar sequence at a nearby replication site to re-engage the stalled complex, leading either to a deletion or duplication of the circumvented segment of DNA [63]. Microhomology-mediated break-induced repair is another mechanism believed to give rise to CNVs under stressed cellular conditions in which traditional break-induced repair does not occur and therefore homologous sequences are identified to continue replication [57,64]. Overall, there are numerous pathways and processes that can lead to the formation of a CNV.

Identification of CNVs across the genome has proven to be challenging because of the dependency on probe density to increase resolution and the physical limitation of the number of probes that can be placed on a single array. Array design technology continues to advance and undoubtedly will increase our ability to identify CNVs by enhancing genome resolution via probe density. Two studies in horses have used the EquineSNP70 BeadChip Array to search the equine genome for CNVs [65,66]. The use of the equine SNP array to identify CNVs highlights the usefulness of the SNP array, but, there are limitations when SNP arrays are used solely for the purpose of identifying CNVs. The probes present on SNP arrays are often evenly distributed across the genome, thus spanning large distances and allowing only for the identification of large CNVs. The SNP arrays also are not well suited for identifying CNVs in structurally complex regions (e.g., gene families, segmentally duplicated regions). Probe design often is difficult in these regions, thus they are excluded from the array [67-69].

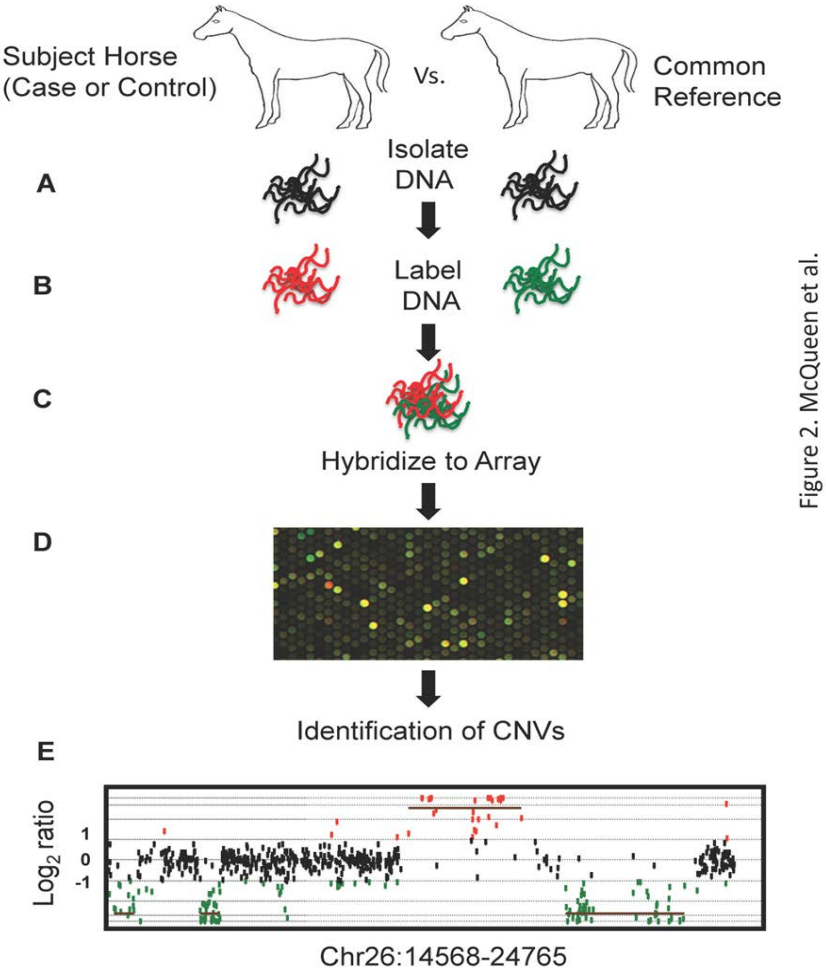


Figure 1. 2. Comparative genomic hybridization method to identify CNVs in horses. (A) Genomic DNA is isolated from subject horses (cases and controls) and a single reference horse. (B) Genomic DNA from the subject horses are independently labeled with a red dye and genomic DNA from a single reference horse is labeled with a green dye. (C) Labeled DNA from a single subject horse and the reference horse are mixed together at equal ratios and competitively hybridized onto a comparative genomic hybridization array. (D) Fluorescent image of array after hybridization of subject and reference DNA. The spots on the array represent individual oligonucleotides. Yellow spots reflect regions with equal DNA content, and red and green spots reflect unequal ratios of DNA between the subject and reference horse, respectively. (E) Plot of normalized log2 ratios of oligonucleotides on the array. The Y-axis represents normalized log2 ratios of fluorescent signals for each spot on the array. The X-axis represents the relative genomic coordinates of each oligonucleotide. For example, a log 2 ratio < 1 and > -1 (black dots) indicates equal DNA content between the subject and reference horses. A log2 ratio >1 and < -1 indicates unequal DNA content between the subject and reference horses.

Several studies have used technologies other than SNP arrays to identify CNVs in horses, such as next generation sequencing (NGS) and comparative genomic hybridization arrays (aCGH) [52,70-74]. Arrays for CGH are designed by tiling oligonucleotide probes across the genome to which DNA of interest then can be hybridized for identification of CNVs (Figure 1.2). Use of aCGH also has limitations for identification of CNVs, principally related to probe placement and density. The currently published equine arrays are a whole genome tiling array (*i.e.*, probes tiled across the whole genome) and an exon tiling array (*i.e.,* probes tiled across non-coding and coding exons of genes) [71,72]. Thus, these arrays only permit evaluation of CNVs within specific regions of the genome. The results of studies identifying CNVs by NGS are limited by variation in read-depth (*i.e.,* the number of copies of sequences aligned to a specific area) across the genome and the size of the CNVs identified. Specifically, CNVs of lengths ranging from 197 bp to 3.5 Mb have been identified and confirmed using a CGH array designed to identify CNVs in genes of the equine genome [71]. In a subsequent study using NGS, CNVs ranging in length from 3.74 kb to 4.84 Mb were identified [70]. There is, however, a trade-off when using either approach. Targeted arrays can identify smaller CNVs, but they are only able to identify CNVs within regions targeted on the array. Conversely, NGS can be used to identify CNVs throughout the entire genome, but NGS approaches to identifying CNVs are limited because of their bias towards larger CNV size. Although there are discrepancies among the approaches used to identify CNVs, the studies to date have identified CNVs in genes involved in similar pathways, such as sensory perception, signal transduction, and immune related

pathways. Results from some CNV studies of horses also have found concordant results between aCGH and NGS whole genome sequencing in which CNVs of horses have been shown to be enriched in genes relating to sensory perception, signal transduction, and immune related functions [71,72].

Our laboratory conducted a CNV-based GWAS by applying the aforementioned equine exon tilling array [71] to the 72 foals studied in our SNP GWAS [52]. Although similar lengths and numbers of CNVs were observed in these foals as in the previous report using this array, no CNVs were significantly associated with *R. equi* pneumonia in these foals. This finding does not preclude the possibility that CNVs contribute to susceptibility to *R. equi* pneumonia because only CNVs within exons were considered. The CNVs located within other elements such as promoters and silencers that were not detected by the array might influence the odds of foals developing *R. equi* pneumonia. Moreover, sample size was small, which might have limited our ability to detect anything less than very strong associations. Future efforts should include the design and implementation of adequately powered studies using tiling arrays focused at gene promotors, gene expression enhancers, and other regulatory elements that are both near and within genes. Much remains to be investigated to characterize CNVs in horses and to study the role of CNVs in susceptibility to *R. equi* foal pneumonia and other diseases of horses. Because CNVs represent a change in genetic content (*i.e.,* deletions and duplications), they may have the potential to greatly impact many phenotypes. A 4.6-kb duplication in an intron has been associated with graying and melanomas in horses [75]. A

16.1-kb duplication has been shown to cause wrinkling of the skin in Shar-Pei dogs [76]. In humans, CNVs are believed to play critical roles in neurodevelopmental disorders, psychiatric disorders, and cancers [77,78]. A number of studies have described CNVs in cattle. Overall, the CNVs identified to date are enriched in genes related to immune function and sensory perception, which also has been observed in horses [79].

Next generation sequencing techniques

The invention of next generation sequencing (NGS) technologies has opened a new world of opportunities for understanding genetic variation and its role in disease pathogenesis. The NGS technology has enabled rapid sequencing of the genomes of individuals at a low cost and with maximum genome coverage [40]. Before NGS, the gold standard for sequencing technology was automated Sanger sequencing [80]. Sanger sequencing technology was developed in the late 1970s and later automated to increase throughput [81]. Next generation sequencing technologies differ among companies, but they all share a principal advantage over Sanger sequencing in that they are capable of sequencing multiple DNA fragments (*i.e.,* an entire genome) in a single sequencing reaction (versus sequencing small fragments piece-by-piece in multiple reactions) [82]. The opportunities provided by NGS technology are accompanied by the challenge of managing and analyzing datasets of enormous size. The ability of NGS to generate data has out-paced the ability of scientists to interpret it. Developments in bioinformatic and biostatistical software have facilitated our ability to visualize and make inferences from large datasets.

Both DNA and RNA can be sequenced using NGS methods. Sequencing the genome (DNA) and transcriptome (RNA) offers 2 interrelated but distinct biological approaches. Genome sequencing using NGS can characterize all of the variants known to exist in gene sequences, including single base changes (SNPs), insertions and deletions, CNVs, and genetic variation in non-genic regions. The first application of NGS for genome sequencing in horses yielded the genome sequence of a Quarter Horse mare [70].

Sequencing of the genome, however, does not reflect which elements of the DNA are transcribed. Moreover, transcription generally should be considered at the level of a specific tissue or cell type because of inter-cellular variation in gene expression. Although the DNA sequence is common to all cells in an individual, the genes expressed vary among cells or organs of the same individual. Sequencing RNA yields a snapshot of the expressed genes of the tissue or cell type that cannot be identified by DNA sequencing. The process of sequencing RNA using NGS methods is termed RNA-Seq; it may be applied either to total RNA (all forms of RNA) or specific types of RNA. Most commonly, RNA-Seq is applied to messenger RNA (mRNA) to reflect which portions of the genome are being transcribed in the specimen. Arriving at RNA-sequencing is a multi-step process which first requires deciding from which tissue or body-fluid RNA should be extracted to best answer the biological questions being asked. Briefly, isolated RNA is converted to complementary DNA (cDNA) in order to construct a library that represents all of the RNA isolated and to be sequenced (Figure 1.3). The representative libraries then are sequenced, generally in a paired-end fashion. Paired-

end sequencing reads are generated by sequencing from both ends of a cDNA fragment (*i.e.*, from the 5' end of both strands of the cDNA fragment). Paired-end reads are extremely valuable because 2 complementary pieces of information have been generated about the same cDNA fragment, and this greatly increases the accuracy of mapping these RNA sequences back to their respective genes of the genome.

RNA-Sequencing is an invaluable tool for gaining insight into biologically relevant questions such as differences in gene expression by different alleles and gene expression of a target specimen under different biological or biochemical conditions. Several downstream RNA-Seq

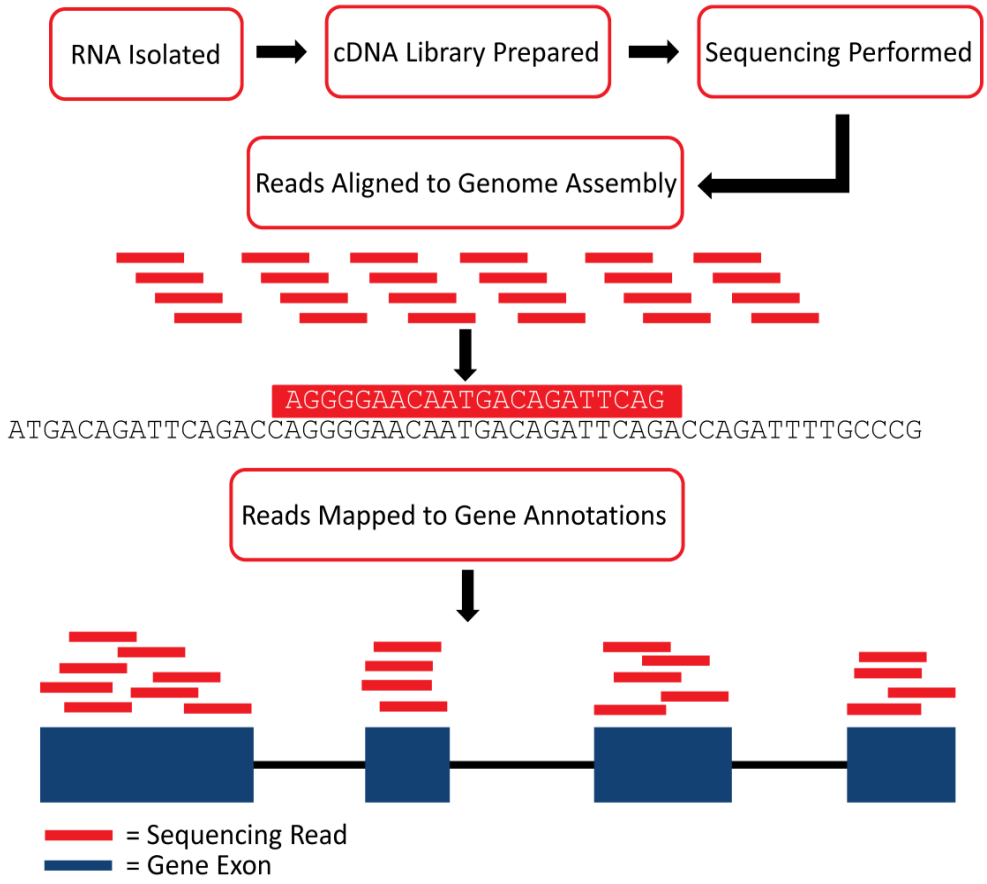


Figure 1.3. RNA-Seq flowchart. Isolated RNA is converted to cDNA, a stable molecule, which can then be amplified and sequenced. Sequencing reads are then aligned to the genome assembly (sequence only) to identify their locations based on nucleotide matches. Mapping the reads to a gene annotation list will generate the number of sequencing reads that have aligned with a particular gene and are called counts. These counts at any particular gene are representative of the amount of gene expression in the sample and can be compared across horses to identify differentially expressed genes.

processing and analysis programs can be used to identify differentially expressed genes, novel transcripts, and multiple isoforms of gene transcripts in order to find answers to biological or clinical questions [83-85]. The conclusions inferred from these analyses can lead to identifying potential biological pathways and processes that can be targeted for development of novel interventions, including treatments and preventative measures. Several studies have reported the application of RNA-Seq in horses in attempt to identify differentially expressed genes [86-95]. To the authors' knowledge, the first report of RNA-Seq in horses was an effort to characterize the transcriptome and tissue-specific expression profiles from 8 equine tissues [94]. A subsequent study focused on characterizing gene expression by RNA-Seq in immunologically active tissues [95]. Investigators have used RNA-Seq to characterize the expression profile of genes critical to the differentiation and regulation of cells during embryogenesis [90], and to characterize the expression and inferred function of RNAs in the equine sperm transcriptome [89]. Several studies also have used RNA-Seq in horses to identify differentially expressed genes when comparing blood, muscle (obtained by biopsy), or both before and after exercise or racing [86,87,92]. These studies have successfully identified pathways involved in stress during and while recovering from exercise. Others studies have sought to answer more specific question such as identifying expression differences in the cartilage of the metacarpophalangeal joints of young and old horses in an attempt to shed light on genes involved in the development and aging of cartilage [93]. Use of RNA-Seq of hoof lamellar basal epithelial cells has been performed to identify cell-signaling pathways indicative of the early stages of laminitis [91]. Using

RNA-Seq, an association has been demonstrated of a long terminal repeat (a genetic element inserted in the past by conversion of viral RNA to cDNA and subsequently incorporated in the genome of the host) with congenital stationary night blindness and leopard spotting in horses [88].

Our laboratory currently is analyzing RNA-Seq data to identify differentially expressed genes of foals representing the 3 genotypes of the *TRPM2* SNP identified in our SNP-based GWAS to better understand the role of this (and possibly other) gene in susceptibility to *R. equi* pneumonia. We also currently are applying RNA-Seq to leukocytes collected from healthy and *R. equi*-affected foals to gain insights about gene expression of these immune-related cells. These studies will further our understanding of *R. equi* pathogenesis and, hopefully identify critical biological pathways and processes involved in disease development.

The genetic basis of a common and complex disease such as *R. equi* pneumonia is likely polygenic. Gene expression profiling by RNA-Seq thus will be an essential step in understanding the relationships and interactions of multiple genes with this disease. The identification of genes that are up- or down-regulated after pathogen exposure can reveal host responses critical for defense against infection. When evidence of differential gene expression is identified by RNA-Seq (or other methods), it then becomes necessary to understand the mechanistic cause driving the change in expression (*i.e.,* variation within

regulatory elements, changes in epigenetic modifications, structural variation, post-transcriptional and post-translational modifications).

Conclusions

Research findings regarding genetic relationships with disease continue to substantiate that most common and complex diseases are not monogenic. This likely is true for *R. equi* pneumonia. The evidence to date, as summarized in this review, indicates that susceptibility to *R. equi* pneumonia is not controlled by a single gene. It is increasingly clear that both innate and adaptive immune responses as well as their interactions are critical for protecting foals against *R. equi* infection. Genetic association studies have specifically implicated innate immune responses, but innate immune responses are critical for orchestrating adaptive immune responses and it may be an over-simplification to dichotomize these responses. It is likely that there also are epigenetic factors involved in regulating gene transcription of critical immune-related genes, which adds further complexity to the pathogenesis of *R. equi* pneumonia in foals. Future proteomic studies also will be required to follow-up on promising genetic findings as protein concentrations, structures, and interactions are critical to disease development [96]. Proteomic studies may be able to answer critical questions such as protein concentrations in diseased and non-diseased foals and variable consequences related to protein concentrations and their interactions, which cannot be answered with molecular genetic techniques and sequencing. With more genotypic-phenotypic associations being identified in horses, it will be challenging to investigate the causal implications of

genetic variants with functional assays. Mechanistic studies (*e.g.,* knock-out or knock-in genes) can be become very expensive and would not be feasible in horses. Developing rodent models of important equine diseases and use of mechanistic studies in cell culture assay will be required to understand the functional consequences of identified associations with genetic markers. Moreover, it will be important to remain mindful of the agent-related and environmental factors that contribute to disease development. No single genetic tool or technique will identify the factors that render some foals susceptible to *R. equi* whereas others in the same environment remain clinically unaffected. The future will require a multifaceted approach to integration and analysis of data from multiple sources to successfully identify the critical pathways and processes. We believe that molecular genetic and epigenetic methods will play an important role in solving the complex riddle of susceptibility to *R. equi* pneumonia in foals.

CHAPTER II

IDENTIFICATION OF GENOMIC LOCI ASSOCIATED WITH *RHODOCOCCUS EQUI* SUSCEPTIBILIY IN FOALS*

Introduction

Rhodococcus equi is an important intracellular pathogen affecting horses, most commonly among foals in which it causes chronic, suppurative bronchopneumonia [29], as well as extrapulmonary disorders [97]. The cumulative incidence of pneumonia caused by *R. equi* may be high at breeding farms with affected foals, and this disease may adversely impact future racing performance [98]. At affected farms, a varying proportion of foals will develop clinical signs of pneumonia while the other foals remain free of the disease; however, subclinical pneumonia can occur following either experimental or natural infection with *R. equi* [99-101].

Although the factors contributing to *R. equi* pneumonia are complex, recent evidence suggests that some horses may be genetically predisposed to this condition [19,29,30]. Identifying the genetic and biological basis of susceptibility, or perhaps resistance, to *R. equi* pneumonia in foals is important, because it might lead to the development of diagnostic and therapeutic tools to manage at-risk foals on breeding farms and might shed light on critical host defense mechanisms. Currently, single nucleotide

*Reprinted with permission from "Identification of Genomic Loci Associated With *Rhododococcus Equi* Susceptibility in Foals" by McQueen et al., 2014. *PLos ONE*, 9:e98710, 2014 by CM McQueen.

polymorphism (SNP)-based genotyping platforms are available for performing genome-wide association studies (GWAS) in horses [102]. Use of SNP-based genotyping platforms to identify genomic regions associated with particular phenotypes in animals is growing at a rapid pace [45,46,102-105]. As a result, researchers, veterinarians, and producers increasingly rely on data from these studies to make important production and management decisions [106,107].

Although high-density SNP arrays are powerful tools for performing association studies, they are often inadequate for examining structurally complex regions, particularly those enriched with copy number variants (CNVs) [67]. Results from the 1000 Genomes project estimate approximately 20% of CNVs are not in linkage disequilibrium with flanking or tagging SNPs [68], indicating that additional testing is required to accurately genotype these variants. The identification of CNVs is further complicated by the probe placement and design of most commercial SNP arrays [67-69]. Comparative genomic hybridization (CGH) arrays are optimized for genotyping CNVs. Using SNP and CGH arrays together may, in some instances, increase the power of a GWAS by expanding the number of informative markers, particularly within structurally complex regions of the genome [69].

In horses, CNVs are present in genes involved in many biological processes and may underlie or modify many common and disease traits [65,66,70]. Of the CNVs in horses identified to date, most are enriched in genes involved in sensory perception, signal

transduction, and metabolism [70]. In other animal species, including horses, CNVs often affect genes regulating the immune system, particularly the MHC; they may also be causative or modifying variants of many immune related conditions [108-115].

The genetic basis of susceptibility or resistance to *R. equi* pneumonia has not been explored on a genome-wide basis. Here, we describe independent SNP- and CNV-based GWAS to identify genomic loci associated with *R. equi* pneumonia in Quarter Horse foals. We identified a number of regions associated with *R. equi* pneumonia, including a region on chromosome 26. Located within this region is the transient receptor potential cation channel, subfamily M, member 2 (*TRPM2*) gene that encodes a protein associated with neutrophil function.

Materials and methods

Ethics statement

All protocols for this study were reviewed and approved by the Clinical Research Review Committee (CRRC Protocol 10-12), College of Veterinary Medicine & Biomedical Sciences, Texas A&M University. This study was carried out on private land (33°37′14″N 100°19′22″W) and specific permissions for use were granted by GPB. During the time this study was conducted, research involving client-owned animals at Texas A&M University was not subject to review by the Institutional Animal Care and Use Committee. Written informed consent for participation was obtained for all foals

included in the study, and the 6666 Ranch provided access to the foals included in this project. This study did not involve any endangered or protected species.

Study population

The 6666 Ranch was selected as the site for this study because it agreed to provide access to foals, had history of *R. equi* pneumonia among foals with a cumulative incidence of ≥ 15% for the preceding 3 years, and because the farm's veterinarian/ general manager (GPB) was conducting a separate study during 2011 evaluating screening tests for *R. equi* pneumonia in foals, which was directed by one of the authors (MKC). The screening test evaluation required that treatment was not initiated for any foal on the basis of screening test results alone, and that the veterinarians making decisions about diagnosis and treatment of *R. equi* pneumonia were not informed of the results of screening tests. Each foal at the farm underwent bilateral thoracic ultrasonographic examination at 2-week intervals, beginning at 3 weeks of age either until 19 weeks of age or until the foal developed clinical signs of pneumonia (as described below). Ultrasonographic examinations were performed by a veterinarian who did not participate in diagnosis or treatment of *R. equi* pneumonia. The anatomic location (left versus right hemithorax; intercostal space; and, dorsal, middle, or ventral region) and maximal diameter of any areas of pulmonary abscesses or consolidation were recorded. In addition, the total number of lesions was counted.

All foals [N = 248] born at the farm during 2011 were eligible to be included in the study. All foals were monitored daily by farm personnel for clinical signs of pneumonia until 20 weeks of age. Clinical signs suggestive of pneumonia included fever, lethargy, signs of depressed attitude, cough, nasal discharge, polysynovitis, tachypnea, increased respiratory effort, respiratory distress, and detection of a tracheal rattle or pulmonary crackles or wheezes via thoracic auscultation. For each foal that developed clinical signs of pneumonia, thoracic ultrasonography and collection of a trans-endoscopic tracheobronchial aspirate (TBA) sample with a commercially available triple-guarded catheter (Triple stage tracheal wash catheter, MILA International Inc., Erlanger, KY) were performed. Between uses, the endoscope was disinfected with a 3.4% glutaraldehyde solution (CIDEX-PLUS, Advanced Sterilization Products, Irvine, CA) following a standard protocol used in our laboratory and known to be microbicidal against *R. equi.* Each sample of TBA fluid was submitted for microbiologic culture and cytologic evaluation to the Texas Veterinary Medical Diagnostic Laboratory in College Station, Texas.

Foals with *R. equi* pneumonia (clinical group; N = 43 [17%]) were defined as those having signs of pneumonia at 3 to 20 weeks of age, ultrasonographic evidence of peripheral pulmonary consolidation or abscesses at the time of examination for clinical signs of pneumonia, and *R. equi* detected in TBA fluid via microbiologic culture, and cytological evidence of gram-positive intracellular coccobacilli in the TBA sample. Subclinical foals (N = 156 [63%]) were defined as those having ultrasonographic

evidence of peripheral pulmonary consolidation or abscesses, but lacking clinical signs of pneumonia [116]. Unaffected foals (N = 49 [20%]) were classified as having no clinical signs of pneumonia and no ultrasonographic evidence of pulmonary consolidation or abscessation. From each of the 3 groups of foals (i.e., clinical, subclinical, and unaffected), 24 foals were selected randomly for the SNP- and CNV-based genome-wide association studies. The rationale for including 24 foals was based on funding available to conduct the study (rather than an *a priori* sample size calculation).

DNA samples and isolation

A blood sample (4 mL) was collected by jugular venipuncture into a tube containing acid citrate dextrose (ACD) as an anticoagulant from the first (i.e., age 3 weeks) blood sample obtained from each foal. Genomic DNA was isolated using a standard phenol-chloroform isoamyl extraction protocol from these blood samples from each foal [70].

SNP genotyping and data analysis

The SNP genotyping was performed at Gene Seek (Neogen, Lincoln, NE) using the EquineSNP70 BeadChip Array (Illumina, San Diego, CA). The resulting SNP genotypes were analyzed using the PLINK analysis package [117]. Genotypes were determined for each animal and then filtered (i.e., excluded) on the basis of missingness per individual (> 10%), missingness per SNP (> 10%) minor allele frequency (< 5%), and absence of Hardy-Weinberg equilibrium (P< 0.001), as described by

Raudsepp *et al.* [45]. A standard chi-square association test (Max(T) permutations [N = 10,000]) based on a binary outcome of disease status using a case-control design was performed using PLINK [44,45]. Genotype ped files were loaded into PLINK and foals were assigned a phenotypic status of either affected or unaffected (case/control). A P value of $P < 1 \times 10^{-5}$ was considered evidence of association [118]. Population stratification was determined using plots of the observed versus the expected $-\log_{10}$ P values of Cochran-Armitage trend tests and by determining the genomic inflation factor, λ, using the R package *GenABEL* [119]. Using the R package pedigreemm [120], mixed-effects logistic regression with sire modeled as a random effect was used for the association test in comparisons showing evidence of population stratification; SNPs with any genotype represented fewer than 10 times were removed from analysis to permit model convergence [119-121]. All SNP array data have been deposited in NCBI's Gene Expression Omnibus (GEO) [122] and are accessible through GEO Series accession number GSE57510 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57510).

Joint analysis

Genotyping for the joint analysis was performed using a tetra-primer AMRS PCR genotyping reaction [123] of an individual SNP (SNP ID:UKUL3936) present on the EquineSNP70 array and located within the *TRPM2* gene (Forward outer: 5`-ATCAGCCAGACACTCCAGGCATGACAT-3`; Forward inner: 5`-CATCCTCCTCAGCCACCTGCATCTTTT-3`; Reverse outer: 5`-ATCTCAGAAGGAGCTGCCATGCCTACC-3`; and, Reverse inner: 5`-

GTATCTTCAGGACCACCCTCCTGACGC-3`). The primers were designed using Primer3 software [124] and synthesized by Sigma-Aldrich (St. Louis, MO). The PCR reactions were performed under these conditions: 9.8 µl mili-Q H_20, 4 µl Taq FlexiBuffer (Promega, Madison, WI), 2 µl $MgCl_2$, 0.4 µl dNTPs, 0.1 µl Taq (Promega), 1 µl forward and reverse inner primer, 0.1 µl forward and reverse outer primer, and 1.5 µl of DNA at 50 µg/µl. Cycling conditions were as follows: 94°C for 2 m; 35 cycles at 94°C for 1 m; 65.8°C for 1 m; 72°C for 1 m; and, a final extension at 72°C for 2 m. PCR amplicons were resolved on a 2% agarose gel.

Genotype data from the joint analysis were analyzed using logistic regression on the basis of the binary outcome of disease (pneumonia versus each of the respective comparison groups [i.e., clinical foals, subclinical foals, and unaffected foals]). The association of disease with genotype for the SNP was expressed as the odds ratio (OR), estimated from logistic regression modeling; 95% confidence intervals were estimated using maximum likelihood methods. Models were fit for comparisons of clinical versus subclinical foals, clinical versus the combination of subclinical and unaffected foals, and clinical versus unaffected foals. Models were fit using S-PLUS statistical software (Version 8.2, TIBCO, Inc., Seattle, WA). A significance level of $P < 0.05$ was used for the analyses.

CNV detection and analyses

Copy number variants were identified using a previously reported equine exome array for comparative genomic hybridization (CGH)[71]. An individual Quarter Horse mare was used as the reference sample for each CGH experiment [70]. Array CGH was performed using methods described by Doan *et al.* [71]. Briefly, genomic DNA was sonicated and then labeled with the Cy5 (experimental) and Cy3 (reference) AlexaFluor dyes using the BioPrime Plus labeling kit (Invitrogen, Carlsbad, CA). Two µg of reference and experimental DNA were hybridized onto the arrays (Agilent Technologies, Santa Clara, CA). The arrays were scanned using an Agilent DNA Microarray Scanner (2-µm settings and 0.05 XDR). Fluorescent intensity values were calculated using Agilent's Feature Extraction 10.5 software (Agilent). Copy number variants, including their corresponding $\log_2$ ratios, were identified using Agilent's Genomics Workbench 7. Copy number variants were called using the ADM-2 algorithm and the following filters: minimum probe span ≥ 3 and average $\log_2$ ratio ≥ 0.5, removal of probes with ≥ 3 standard deviations above or below the mean $\log_2$ fluorescent intensity.

Logistic regression modeling was used to perform 2 separate CNV-based GWAS. The first approach modeled the association of the binary outcome of 2 groups (e.g., clinical versus subclinical) with the $\log_2$ ratio of intensity values (a continuous variable) for each CNV. The second approach modeled the association of the binary outcome of 2 groups with the presence or absence of a CNV (a binary categorical variable) within a CNV region. For the first approach, CNV regions (CNVRs) were determined for the foals

examined. The CNVRs were then filtered to include only CNVs identified in at least 3 foals. The log_2 ratios of probes within each CNVR were then averaged to calculate a single log_2 (CNVR-log_2 ratio) for each CNVR for each foal. The CNVR-log_2 values were used in a logistic regression model to identify associated CNVs among the pairwise comparisons of the 3 groups (case-control design described above). For the second approach, a logistic regression model involving the binary outcome of presence (or absence) of a CNV within each CNVR was used to identify associated CNVs among the pairwise comparisons of the 3 groups. For both approaches, the generated P values from linear modeling or regression analyses were corrected for multiple comparisons using the method outlined by Hochberg *et al.* [125]. Statistical analyses were performed using R (Version 3.0.1; R Statistical Project). All CNV data have been deposited in NCBI's Gene Expression Omnibus (GEO) [122] and are accessible through GEO Series accession number GSE57510 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57510).

Results

SNP-based GWAS

Three case-control GWAS were performed among the 24 randomly selected foals representing each group (clinical group, subclinical group, and unaffected group; Figure 2.1). The number of SNPs excluded on the basis of missingness per individual,

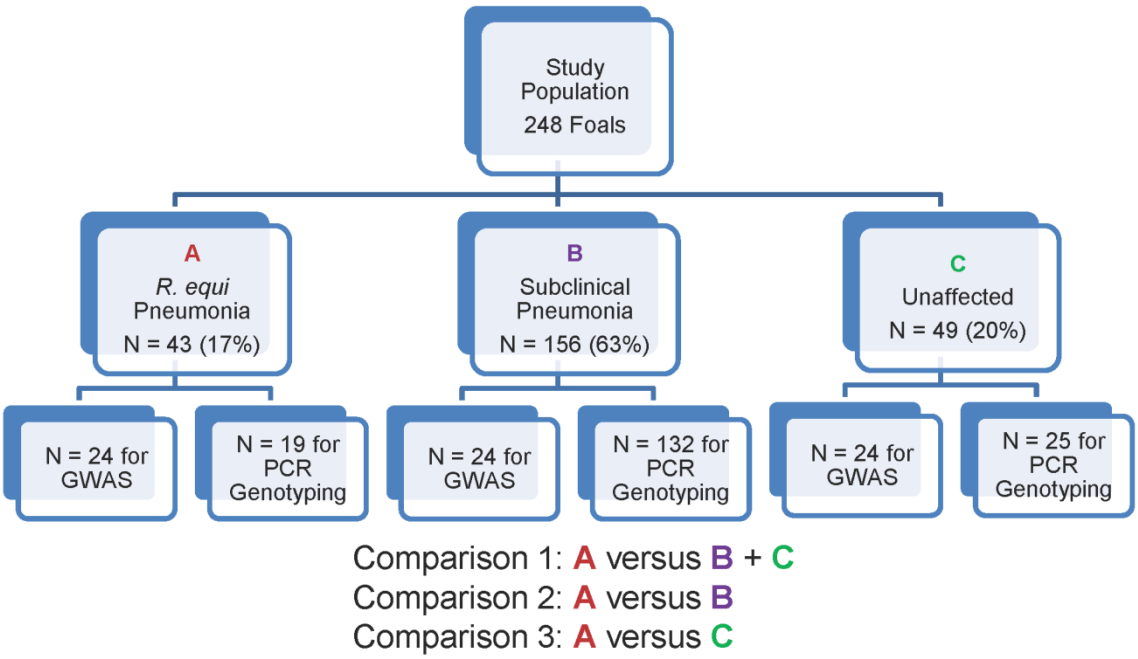


Figure 2.1. GWAS study population. Schematic diagram representing the distribution of the total population into the 3 subgroups (*R. equi* pneumonia foals [clinical], subclinical foals, and unaffected foals), and by genome-wide association studies versus PCR genotyping for TRPM2 SNP. The 3 comparisons among groups are also summarized.

missingness per SNP, and minor allele frequency were 0, 1,292, and 11,157, respectively (12,449 total SNPs). For comparisons 1, 2, and 3 (Figure 2.1), the number of SNPs excluded on the basis of Hardy-Weinberg equilibrium were 0, 83, and 54, respectively. After filtering, the total genotyping rate of the foals was estimated at 99.4%. Comparison 1 (clinical [N = 24] vs. subclinical + unaffected [N = 48]; $\lambda = 1.16$) identified 7 regions showing evidence of moderate association with clinical pneumonia ($P < 1 \times 10^{-5}$) (Figure 2.2 A and Table 2.1). Comparison 2 (clinical [N = 24] vs. subclinical [N = 24]; $\lambda = 1.00$ [Supplementary Figure 2.1 A]) identified 10 regions with moderate association (Figure 2.2 B and Table 2.1). The region associated with clinical pneumonia had a (point-wise) value of EMP1 ≤ 0.0002. Comparison 3 (clinical [N = 24] vs. unaffected [N = 24]; $\lambda = 1.44$ [Supplementary Figure 2.1 B]) identified 2 regions with moderate association (Figure 2.2 C and Table 2.1). Results from each GWAS comparison are provided in Supplementary Table 1.

The λ value of comparison 3 (1.44) suggested evidence of confounding population structure from 1 of the groups, thus mixed-effects modeling with sire as the random effect term was used as an additional association test. There were 23,318 SNPs filtered due to failure to converge in mixed modeling, leaving 40,843 SNPs for evaluation in comparison 3. The logistic mixed-effects modeling reduced the λ from 1.44 to 1.10 (Supplementary Figure 2.1 D), and inspection of the observed versus expected P value plot indicated that none of the smallest P values [highest $-\log_{10}$ P values] observed were greater than expected (Figure 2.2 D and Supplementary Table 2.1). The mixed-effects

model analysis identified 1 region with modest evidence of association (BIEC2_284540; chr15:7,394,044; $P = 9.0 \times 10^{-5}$). The λ values of each comparison suggested the population structure was attributable to foals in the unaffected group because the magnitude of λ was greatest for the comparison of clinical versus unaffected groups (i.e., strongest evidence of population stratification), was 1.00 for clinical versus subclinical groups (i.e., absence of population stratification), and 1.14 when the clinical group was compared with the combination of healthy and subclinical (indicating the healthy group contributed to evidence of population stratification for this comparison).

A region on chromosome 26 (chr26:39,640,172-39,867,963) showed evidence of association with clinical pneumonia (Comparisons 1 and 2), with the strongest evidence of association in comparison 2. Three of the 4 SNPs identified in this region were BIEC2_732054, BIEC2_696979, and BIEC2_696992, with the first lying in a keratin-associated protein (*KRTAP)* gene and the remaining 2 SNPs lying in non-genic locations. The region also contained a potential candidate gene (*TRPM2*) based on biological

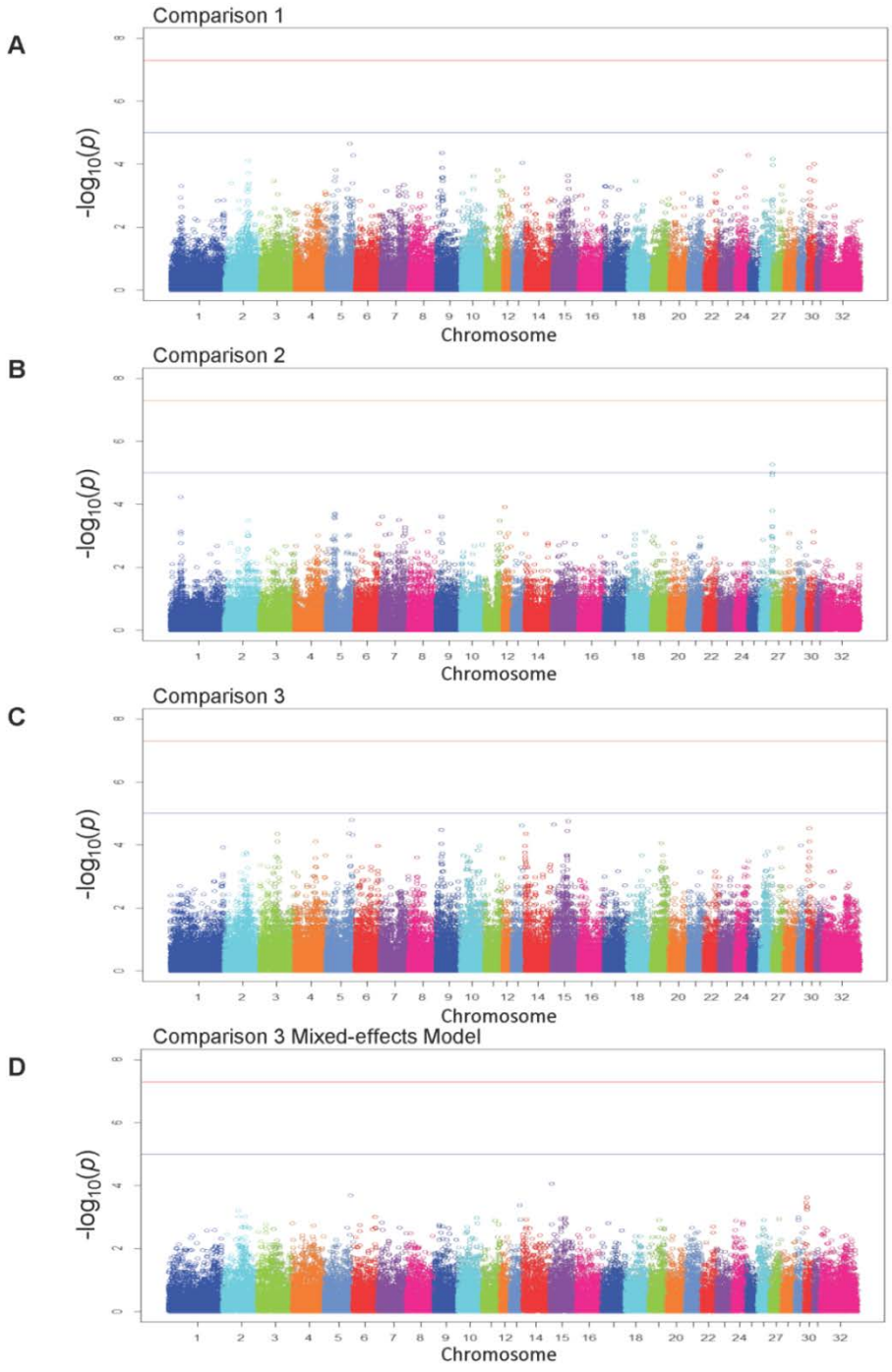


Figure 2.2. GWAS Manhattan plots. Manhattan plots of standard chi-squared significance values for the 3 genome-wide association studies. Manhattan plots for (A) comparison 1, (B) comparison 2, and (C) comparison 3; (D) Mixed effects-model analysis of comparison 3.

Table 2.1. Moderately associated SNPs for each genome-wide association study.

Comparison 1			
SNP ID	**Chromosomal location**	**P Value**	**Odds Ratio**
BIEC2_921509	5:79290169	2.21E-05	0.2
BIEC2_1137861	9:19578673	4.42E-05	9.556
BIEC2_976727	5:91479639	5.10E-05	0.09677
BIEC2_650473	24:41033593	5.16E-05	0.04274
BIEC2_732054	26:39777632	6.78E-05	12.05
BIEC2_696979	26:39861109	6.78E-05	12.05
BIEC2_696992	26:39867963	6.78E-05	12.05
BIEC2_492054	2:80197359	7.73E-05	0.2036
BIEC2_240596	13:31572932	8.94E-05	0.156
BIEC2_825623	30:20687203	9.79E-05	0.2369
BIEC2_867975	30:20690855	9.79E-05	0.2369
Comparison 2			
SNP ID	**Chromosomal location**	**P Value**	**Odds Ratio**
BIEC2_732054	26:39777632	5.54E-06	17.89
UKUL3936	26:39640172	9.93E-06	12.69
BIEC2_696979	26:39861109	1.24E-05	16.43
BIEC2_696992	26:39867963	1.24E-05	16.43
BIEC2_16162	1:35593058	5.80E-05	13.8
Comparison 3			
SNP ID	**Chromosomal location**	**P Value**	**Odds Ratio**
BIEC2_927543	5:88617368	1.63E-05	0.1486
BIEC2_311792	15:54419913	1.75E-05	0.1515
BIEC2_284540	15:7394044	2.21E-05	0.08571
BIEC2_240596	13:31572932	2.41E-05	0.1163
BIEC2_858227	30:6939831	2.99E-05	0.1512
BIEC2_1078504	9:19582539	3.37E-05	0.1314
BIEC2_310113	15:51485703	3.57E-05	0.1074
BIEC2_310214	15:51596762	3.57E-05	0.1074
BIEC2_324522	15:51634407	3.57E-05	0.1074
BIEC2_921509	5:79290169	4.19E-05	0.1667
BIEC2_784299	3:59782881	4.45E-05	0.1445
BIEC2_239002	14:3055273	4.46E-05	0.1696
BIEC2_976312	5:90703095	4.80E-05	0.09333
BIEC2_928796	5:90706671	4.80E-05	0.09333
BIEC2-815626	30:5321760	7.70E-05	0.1169
BIEC2_784331	3:59874438	7.87E-05	5.8
BIEC2_784332	3:59874495	7.87E-05	5.8

Table 2.1. Continued

BIEC2_869756	4:72076243	7.87E-05	5.8
BIEC2_435811	19:33380654	8.85E-05	0.1784
BIEC2_1137861	9:19578673	9.20E-05	10.09

function [54] identified by the SNP marker, UKUL3936 ($P = 9.93 \times 10^{-6}$; OR = 12.7), which was also located within exon 22 of the *TRPM2* gene. On the basis of this finding, a joint analysis was performed to include results of PCR-based genotyping of the UKUL3936 SNP for the samples from the larger remaining population of foals not included in the GWAS study (i.e., clinical [N = 19], subclinical, [N = 132], and unaffected [N = 25]). Prior to performing the joint analysis, 10 of 72 foals previously genotyped on the SNP array were used to validate the genotyping reaction used for the joint analysis; results of SNP array and PCR genotyping agreed for all 10 foals. The joint analysis comparing clinical versus subclinical foals (comparison 2) revealed that foals from the clinical group were approximately 4-fold less likely to have either an AB (1/(0.23) = 4.3; P = 0.0017) or BB genotype (1/(0.28) = 3.6, P = 0.0574), consistent with a dominant model, possibly with partial penetrance (Table 2.2). Considering just the AA genotype relative to the other

Table 2.2**.** Joint analysis of *TRPM2* SNP UKUL3936

Genotype				
Standard Model	**Clinical foals**	**Subclinical foals**	**P value**	**Odds Ratio (95% CI)**
AA	72% (31/43)	41% (64/156)	NA	1 (NA)
AB	23% (10/43)	47% (74/156)	0.0017	0.28 (0.13 to 0.61)
BB	5% (2/43)	12% (18/156)	0.0574	0.23 (0.05 to 1.04)
Dominant Model				
Not AA	28% (12/43)	59% (92/156)	NA	1 (NA)
AA	72% (31/43)	41% (64/156)	0.0006	3.71 (1.78 to 7.76)
Additive Model				
f(A)*	2(0 to 2)	1(0 to 2)	0.0014	2.83 (1.51 to 5.31)
	(72/86) 84%	(202/312) 65%		

*** Median (range) reported for frequency of allele A, along with the proportion of A alleles among all alleles represented for each group. Joint analysis includes genotypes derived from SNP array and PCR genotyping.**

genotypes, the odds of disease were approximately 3.7-fold greater for foals with the AA genotype (Table 2.2; P = 0.0006). Using an additive model, there was a significant (P = 0.0014) association of the A allele in comparisons between the clinical and other groups, with an estimated odds of a clinical classification being increased nearly 3-fold for each copy of the A allele. Examination of the genotype data, however, suggested an additive

model was unlikely: the ORs for the AB and BB types relative to the AA genotype were similar. The AA genotype of the TRPM2 gene also was significantly associated with increased odds of *R. equi* pneumonia when considering the results of comparisons of the clinical foals versus all foals (comparison 1; Supplementary Table 2) and for clinical foals versus healthy foals (comparison 3; Supplementary Table 2.3), irrespective of the genetic model.

CNV-based GWAS

Next, array aCGH was performed to genotype CNVs in the 72 foals examined in the SNP-based GWAS. Two reactions failed to meet the minimum quality scores for CNV detection, so they were excluded from the study (foals 153 and 278). Collectively, 6,727 CNVs were identified among the 70 foals (Table 2.3 and Supplementary Table 2.3). Merging shared CNVs yielded 2,350 CNV regions (CNVR) that were present at 3,492 Ensembl annotated genes (3,442 protein-coding and 50 RNA-coding). The lengths of CNVs ranged from 197 base-pairs (bp) to 7,229.5 kilo-bp (kb), with a mean length of 97.7 kb, median length of 4.4 kb, and mode length of 960 bp.

The association between disease status (the outcome variable) and individual CNVs (dependent variable) was assessed using logistic regression analysis. Because CNV genotypes reflect

Table 2.3. CNVs identified in each clinical group.

	Clinical foals	Subclinical foals	Unaffected foals	P Value
Total CNVs	84 (29 to 592; 2,654)	57 (29 to 338; 1,823)	77 (28 to 254; 2,250)	0.262
Gains	25 (8 to 395; 1,229)	24 (9 to 187; 770)	30 (10 to 152; 957)	0.334
Losses	41 (15 to 197; 1,425)	36 (16 to 151; 1,053)	40 (16 to 139; 1,293)	0.574
CNVs: median (range; sum); Kruskal-Wallis test				

Table 2.4. Top 5 CNV regions identified using logistic regression for the association of *R. equi* with either the binary variable presence or absence of a CNV identified in the region (Presence column) or the $\log_2$ ratio of intensity values of the CNVs (Intensity columns).

Presence		**Intensity ($\log_2$)**	
CNVR	**Location**	**CNVR**	**Location**
232	chr1:158563545-160296515	179	chr1:135750728-135756968
237	chr1:161277315-161277884	391	chr2:88754970-89061174
269	chr2:6304157-6378969	458	chr3:36530104-36530453
536	chr4:21281685-23733388	645	chr5:40060212-40060524
735	chr6:42263179-42282649	753	chr6:59580727-59585139
CNVR identification numbers are provided in Supplementary Table 1.			

differences in DNA content between 2 individuals and are expressed as normalized $\log_2$ ratios representative of varying degrees of copy number gains and losses, separate logistic regression modeling was performed using CNVs as the dependent variable as either 1) continuous variables representing average $\log_2$ ratios of CNVRs or 2) a binary variable representing the presence or absence of a given CNV (see Materials and Methods section). Comparisons among groups of foals were made in the CNV-based GWAS as described above (Figure 2.1). Correction for multiple comparisons revealed no significant ($P < 0.05$) associations of disease with CNVs when considered as continuous $\log_2$ ratios (Table 2.4) or as the binary outcome for presence or absence of a CNV. No association with clinical status grouping was detected on the basis of the total number of CNVs for individuals (Table 2.3).

Discussion

Rhoddococcus equi is an important cause of disease and death in young foals [6]. Multiple factors such as age, environmental conditions including level of exposure to virulent organisms, and genetic background appear to play a role in the occurrence of this complex disease [126-128]. The purpose of this study was to better characterize the genetic basis of susceptibility to *R. equi* pneumonia. One special feature of this study was the phenotypic characterization of foals into those that remained unaffected (i.e., free of both clinical signs and ultrasonographic evidence of disease) through weaning, those that had subclinical pneumonia (i.e., absence of clinical signs but ultrasonographic

evidence of pulmonary lesions), and those that developed *R. equi* pneumonia. At most farms where screening is performed, foals with evidence of subclinical pneumonia receive treatment or other interventions that precludes one from differentiating foals that would have progressed to clinical disease from those that would have remained subclinical. At the farm described in this report, however, foals with subclinical pneumonia were not treated or otherwise managed differently than unaffected foals providing us with the exceptional opportunity to conduct GWAS's with 3 clinically important phenotypes.

A SNP-based GWAS revealed a region on chromosome 26 (chr26:39640172-39867963) that was positively associated with disease. This region contains the *TRPM2* gene, which is associated with neutrophil function. Although, *TRPM2* is an ideal candidate based on its known biological function, an adjacent SNP located within the *KRTAP* gene was more strongly associated than the SNP located with *TRPM2*. Further, investigation of this region (e.g., fine-mapping of the region) is needed before any conclusion can be made regarding the role of *TRPM2* in *R. equi* pneumonia. Nevertheless, *TRPM2* is of interest because it has been demonstrated in mice to play a role in neutrophil-mediated tissue damage [54]. Neutrophils have been shown to play an important role in the outcome of *R. equi* infection. Neutrophil-depleted mice had significantly heavier tissue burdens of *R. equi* following experimental infection than non-depleted mice, documenting a protective role for neutrophils [129]. The neutrophil concentrations at 2 and 4 weeks of age were significantly lower among foals that subsequently developed *R. equi* pneumonia

than among age-matched foals that did not develop pneumonia [130]. Similar protective roles for neutrophils have been documented for other intracellular pulmonary pathogens [129,131,132]. Moreover, *R. equi* has an age-dependent distribution (i.e., foals are usually affected and adults are generally resistant to infection), and age-related differences in neutrophil responses to *R. equi* have been documented [133-135]. Although neutrophils play a role in protecting against *R. equi* infection, they also contribute to lung parenchymal damage of this pyogranulomatous disease [6,136]. In a mouse model of ulcerative colitis, the over-abundance of neutrophil invasion into tissue mediated by *TRPM2* expression, led to increased colonocyte death [54]. Thus, variation in *TRPM2* expression could influence the extent to which neutrophil-induced pulmonary damage occurs following infection with *R. equi*, and this variation could be a crucial determinant of the clinical outcome of infection with *R. equi* and the progression from subclinical to clinical pneumonia. If the *TRPM2* allele implicated in our study were associated with increased *TRPM2* expression, it might consequently be associated with greater likelihood of pneumonia development as a result of greater neutrophil invasion. The functional effects of this *TRPM2* genotype are unknown, however. Interestingly, expression of a splice variant of *TRPM2* was demonstrated to inhibit death of several cultured cell lines [137]. Further evaluation of the functional differences in neutrophilic responses among *TRPM2* genotypes is warranted.

A CNV-based GWAS was conducted and revealed no significant association with disease status. Despite the negative findings, these results are of interest with regard to

better characterizing genetic variation in horses and using CNVs for GWAS with disease outcomes. A search of PubMed reveals no other attempt to perform a CNV GWAS via aCGH in horses and to the authors' knowledge no others have been done. Valuable information was gained in terms of the analysis of CNV data generated from a GWAS using aCGH. The identification of CNVs is based on $\log_2$ ratios of intensity signals that are generated between the reference and sample DNA. The results may be interpreted either as presence or absence of a CNV, based on a threshold intensity value [71], or the actual intensity values themselves. As observed in this study, these 2 outcomes for CNV-calling yielded differing results. We propose that utilizing average $\log_2$ ratios across CNV regions is superior because classification of CNVRs by presence or absence does not further characterize a CNV, whereas analysis of the $\log_2$ ratios allows for the identification of whether the CNV involved gains or losses, and for description of the magnitude of the gain or loss.

We failed to identify significant association with candidate genes previously associated with *R. equi* pneumonia in other breeds of foals [19,29,30]. This may have been attributable to differences among populations of foals studied (e.g, breeds) or study methodology (e.g,, case definitions, methods for detecting polymorphisms, etc.). Nevertheless, a commonality among these studies can be found in their identification of genes pertaining to host defenses against infectious pathogens, such as iron transport and innate immune responses. Conceivably, these apparently discrepant findings may converge on critical

biological pathways or processes that influence susceptibility to infection with *R. equi* (and other intracellular pathogens).

This study had a number of limitations. First, it was likely underpowered for both the SNP- and CNV-based GWAS portions of the study: we only had the opportunity to find SNPs or CNVs with large effects. This lack of power is probably why we did not identify highly significant associations in the SNP-based GWAS for 2 of our 3 comparisons (Figure 1), and only moderate significance for the allele identified in comparison 2. Although sample sizes may be calculated for human studies, methods for incorporating crucial determinants of sample size such as the impact of the relatively longer length of linkage disequilibrium in horses relative to humans remain to be defined. Moreover, the cost of GWAS studies and the limited funding available for equine research can be restrictive. It is worth noting that our SNP-based GWAS did provide sufficient power to identify a SNP associated with *R. equi* that was subsequently substantiated by findings of the joint analysis using PCR testing of additional foals from this population for all 3 comparisons. Nevertheless, larger scale studies are indicated. For example, 1 locus on chromosome 15 (in the promoter region of a chemokine) was weakly associated in both the mixed-effects modeling and standard GWAS analyses for comparison 3: this finding could be simply attributable to chance but also could represent an underpowered association that merits further investigation.

A limitation of this study is that it was restricted to foals of a single breed at a single farm. Further studies are indicated to substantiate whether the observed association holds among other Quarter Horse foals at other farms, and among foals of other breeds. Another limitation was that although the mixed modeling for comparison 3 yielded a marked reduction in the estimated value of λ (from 1.44 to 1.10), the mixed modeling value of λ suggested residual confounding from population structure. Graphical analysis of the results of mixed modeling, however, suggested this was not the case: the plot of observed versus expected P values indicated that fewer than expected small P values were identified and more than expected large P values were observed following mixed-effects modeling, and Manhattan plots of the standard and mixed-effects model GWAS for comparison 3 revealed a shift to larger P values following mixed-effects modeling (Supplementary Figure 2.1). Moreover, the linear regression-based estimate of λ had a poor fit because the data were non-linear in the range of the high P values (data not shown). We interpreted these results to indicate that the mixed-effects modeling had largely corrected for population stratification attributable to sire, and that the study was underpowered. As noted, a limitation of this study is that we lack fine-mapping of the region to determine whether the *TRPM2* allele is a causal variant or simply in linkage disequilibrium with another gene or genes that contribute to susceptibility to *R. equi*. Finally, it must be noted that although the estimated OR for the *TRMP2* SNP was relatively large, it is not of sufficient magnitude to be clinically useful for screening purposes.

Despite the aforementioned limitations, this study has identified a region on chromosome 26 that was associated with *R. equi* pneumonia in foals, and the largest scale GWAS in foals reported to date. Furthermore, it extends current knowledge of equine CNVs and analysis of data from GWAS using aCGH in horses.

Conclusions

In conclusion, the results of this study identify a locus and gene potentially involved in the development of *R. equi* pneumonia in foals. Future studies are warranted to substantiate the association of *TRPM2* gene (and related pathways) with *R. equi* pneumonia and to provide fine-mapping of the region on chromosome 26 implicated in this GWAS.

CHAPTER III

USE OF RNA-SEQUENCING TO INVESTIGATE FUNCTIONAL IMPLICATIONS OF *RHODOCOCCUS EUQI* ASSOCIATED GENTOYPES

Introduction

Rhodococcus equi is a pathogen that predominantly affects young foals causing pneumonia as well as extrapulmonary disorders [1,6,53,97,116]. Currently, there is no approved or effective vaccine for protection against *R. equi* pneumonia, and other preventative interventions, such as transfusion of hyperimmune plasma, are expensive, labor-intensive, and incompletely effective [3,6]. Evidence has shown that *R. equi* taken up by the host must bear the virulence associated protein A (VapA) gene in a plasmid in order to cause disease; however, presence of the plasmid and VapA alone is not sufficient to cause disease indicating host factors are of great importance [7,8]. In addition to anecdotal reports of some mares having multiple infected foals while other mares in the same environment never have an affected foal, several candidate gene studies suggest a genetic basis of *R. equi* susceptibility and have been reviewed [19,22,29,30,138]. Due to gene selection biases and phenotypic misclassification, the associations from candidate gene studies have been weak and potentially biased [17].

Recently, our laboratory identified a region on chromosome 26 associated with *R. equi* pneumonia in a genome-wide association study (GWAS)[15] using a commercially available SNP array [39,52]. Four SNPs were associated with clinical disease in a region

spanning several predicted genes. One of the SNPs was well suited to serve as a marker because it was located within the transient receptor potential cation channel, subfamily M, member2, candidate gene (*TRMP2*) and could be easily and accurately genotyped [52]. This SNP was associated with 3 genotypes designated AA, AB, and BB alleles. Because SNPs are merely indicators of location and are very rarely actual causal mutations, it remains unclear which genes in this region might explain the observed association of genotype with disease. More importantly, these SNPs alone do not provide any functional information regarding the relationship between genotype and phenotype across this region. The marker SNP in *TRPM2* is a synonymous substitution and does not change the amino acid or protein sequence. Thus, other approaches such as investigating gene expression are necessary to further investigate the observed association of the *TRPM2* SNP and *R. equi* pneumonia.

Investigating the whole transcriptome (*i.e.*, the entire translated genome) using RNA-Seq provides an unbiased approach for gene expression analysis because all potential genes and their isoforms are considered [139] yielding greater possibility for identifying true expression phenotypes associated with the genotype(s) of interest. Using RNA-Seq enables the characterization of the transcriptional activity across the region of interest (i.e., targeted analysis) as well as across the remainder of the genome (i.e., untargeted analysis), such as differences in total expression of genes, alternatively spliced transcripts, and truncated gene transcripts. Thus, we used RNA-Seq to perform targeted and untargeted associations of gene expression with both genotype and phenotype

previously associated with *R. equi* infection. For our targeted approach, we determined which genes in the region of interest on chromome 26 were expressed differentially by each of the 3 marker genotypes identified for the *TRPM2* SNP (*i.e.*, AA, AB, or BB alleles) as well as to identify associations between gene expression profiles and clinical phenotypes (Supplementary Figure 3.1). In addition to our targeted analysis, we also examined genes differentially expressed across the entire transcriptome to look for other genes that were differentially expressed that might be associated either with genotype or with susceptibility to *R. equi* pneumonia. The objectives of this study were to identify biological pathways and processes involved in susceptibility to *R. equi* pneumonia as well as to characterize the Quarter Horse transcriptome [86,87,94].

Materials and methods

Study population

All protocols for this study were reviewed and approved by the Institutional Animal Care and Use Committee (AUP: IACUC 2013-0259 CA, Approval Date: 1/29/2014). Written informed consent for participation was obtained for all horses included in the study, and the 6666 Ranch provided access to the horses for use in this project. The study population was derived from the source population for our GWAS described above (*viz.*, foals born during 2011 at the 6666 Ranch) [52], which had a cumulative incidence of *R. equi* pneumonia of 17% during 2011. We identified 51 horses still remaining at the 6666 Ranch that had been foals during our previous study in 2011;

horses (formerly foals) of each of the 3 genotypes of interest (*i.e.*, AA, AB, and BB) identified from our GWAS were available for testing. Foals were randomly selected for participation in the RNA-Seq portion of this study which included 3 AA, 4 AB, and 5 BB genotypes. This sample size was determined by the funding available for this project and the cost of RNA-Seq.

Sample collection and RNA-Seq

A whole blood sample (5 mL) was collected by jugular venipuncture into 2 Paxgene RNA Vaccutainer tubes (PreAnalytiX, Hombrechtikon, Switzerland) at the 6666 Ranch, Guthrie, TX, June 25, 2014, to permit the RNA to be stabilized for transport to Texas A&M University. The blood samples were processed with the MagMax Paxgene RNA purification kit (Life Technologies) to isolate total RNA. The Texas A&M Agrilife Genomics and Bioinformatics facility constructed a cDNA library using the TrueSeq RNA preparation kit (Illumina) with a polyA selection step,and performed RNA-Seq. Samples were pooled and sequenced in duplicate (2 lanes) on a HiSeq 2500 (Illumina) to account for any lane biases that might arise. The HiSeq2500 generated 125-base-pair (bp) paired-end (PE) sequencing reads for expression analyses.

Gene expression analysis for genotypic and phenotypic comparisons (targeted and untargeted)

Raw data output of the RNA-Seq was processed by our laboratory using multiple online resources provided by the Texas A&M Institute for Genome Sciences and Society. The generated PE reads were aligned to the entire equine genome reference sequence (Equcab2) using *Tophat* [84]. Aligned reads were then mapped to equCab2 ENSEMBL gene annotation and read counts assigned to each sample at each gene using *HTSeq* [140]. *HTSeq* generated count tables using the intersection nonempty parameter to account for ambiguous reads (*i.e.,* overlapping reads that map to overlapping genes).

Three comparisons were chosen to test for differentially expressed genes within the horses. The first comparison made was between the AA genotyped foals (n = 3) and the non-AA genotyped foals (*i.e.,* the combined total of AB and BB genotyped foals (n = 9)). The AB and BB genotypes were combined on the basis of our previous results indicating that the odds of disease were equivalently greater in the AA homozygotes relative to the AB genotype, the BB genotype, or their combination [52]. The next comparison tested for differential expression among the 3 genotypes (*viz.,* AA, AB, and BB) to identify any genes differentially expressed relative to each genotype. A third comparison tested for differential gene expression between horses that had clinical cases of *R. equi* pneumonia as foals (n = 2) and horses that had not developed clinical signs of pneumonia as foals (i.e, subclinical pneumonia or no pneumonia; n = 10). All analyses

were performed using R statistical software (Version 3.0.1; R Statistical Project), the *edgeR* package, and *cuffdiff* [84,141-145]. All normalizations and calculations were performed using the recommended pipeline per the *edgeR* and *Tophat* authors' instructions. Briefly, the *HTSeq* generated count table was filtered to remove genes to which 0 reads mapped in all samples. In *edgeR* the tabulated read counts were then normalized and tag-wise and common dispersions estimated. Differentially expressed genes were defined as those genes having a false discovery rate (FDR) ≤ 0.05 that were identified with the *exactTest* function. The *cuffdiff* function of the *Tophat* tool suite was used to calculate differentially expressed genes using the default settings with the accepted_hits.bam file from the *Tophat* mapping step used as input. Genes having a q-value (FDR) ≤ 0.05 were considered differentially expressed. *Cuffdiff* was only used to test for differential expression using the AA vs. AB/BB and genotypic pairwise comparisons described above.

Biological pathway analysis was performed using the Ingenuity Pathway Analysis (IPA) (Qiagen, Venlo, Netherlands; www.ingenuity.com) tool-kit. Output files from both *edgeR* and *cuffdiff* were used in their respective pathway analyses. Input files for IPA used the following information from the *edgeR* or *cuffdiff* outputs: a column containing an Ensembl gene identifier, a corresponding RefSeq gene name where applicable, the log fold change of expression for each gene between the 2 groups tested, and a FDR for each gene tested.

Visualization of RNA-Seq read coverage across the region of interest was carried out using *BEDTools* and the UCSC online genome browser by assigning reads from the accepted_hits.bam output files from the previously described *Tophat* mapping step [146,147]. Based on evidence from visualization of the RNA-Seq reads, transcript variation across the region of interest (*viz.,* chr26; targeted analysis) was also assessed. The *cufflinks* function of the *Tophat* tool suite was used to identify alternatively spliced and novel transcripts at the *TRPM2* locus [84]. Briefly, *cufflinks* does not use an annotation to map and calculate read counts at genes, rather it uses a *de novo* approach for calculating transcripts and isoforms prior to mapping reads to an annotation.

Targeted *TRPM2* analysis was performed by using the datasets generated from the *cuffdiff* comparisons previously described. *Cuffdiff* generates output files containing the transcript (gene, encompassing all isoforms) abundances and isoform (alternatively spliced transcripts) abundances which were extracted and expression level compared for specific individual transcripts.

Table 3.1. Transcriptome-wide differentially expressed genes identified by *edgeR* analysis.

Ensembl ID	RefSeq ID	Log FC	Log CPM	P Value	FDR
ENSECAG00000022239	ANKRD22	-5.82	0.23	0.0000	0.0000
ENSECAG00000006492	DQB	-2.00	7.30	0.0000	0.0028
ENSECAG00000006662	MPO	-3.22	-0.44	0.0000	0.0483
ENSECAG00000017147	C15orf52	-3.08	-1.66	0.0000	0.0968
ENSECAG00000002249	PLEKHG4B	-4.64	-1.57	0.0000	0.0983
ENSECAG00000008171	N/A	-3.80	-0.68	0.0000	0.1476
ENSECAG00000008238	S100A5	-3.41	-1.98	0.0001	0.1476
ENSECAG00000016666	OMG	0.89	6.37	0.0001	0.2650
ENSECAG00000006656	N/A	-3.08	-2.89	0.0002	0.4258
ENSECAG00000024043	CSTA	-2.09	2.50	0.0003	0.4904

Table 3.2. Transcriptome-wide differentially expressed genes identified by *cuffdiff* analysis.

Ensembl ID	RefSeq ID	Log_2 FC	P value	Q Value	Significant
ENSECAG00000024043	CSTA	-2.08	0.0001	0.0122	yes
ENSECAG00000024259	DQA	-1.83	0.0001	0.0122	yes
ENSECAG00000019922	ADAMDEC1	-1.39	0.0001	0.0122	yes
ENSECAG00000009142	DQA	-1.34	0.0001	0.0122	yes
ENSECAG00000020816	PDLIM1	-1.32	0.0001	0.0122	yes
ENSECAG00000015109	N/A	-1.28	0.0001	0.0122	yes
ENSECAG00000012883	CEBPE	-1.27	0.0001	0.0122	yes
ENSECAG00000023062	N/A	-1.26	0.0001	0.0122	yes
ENSECAG00000019130	SIRPG	-1.22	0.0001	0.0122	yes
ENSECAG00000011315	EMR3	-1.17	0.0001	0.0122	yes
ENSECAG00000013660	C1orf186	-1.16	0.0002	0.0371	yes
ENSECAG00000025078	SUSD2	-1.05	0.0001	0.0122	yes
ENSECAG00000016730	CSMD1	-0.99	0.0001	0.0122	yes
ENSECAG00000001282	CCR3	-0.95	0.0001	0.0122	yes
ENSECAG00000008322	GZMH	0.79	0.0002	0.0371	yes
ENSECAG00000007621	TRIP11	0.86	0.0003	0.0435	yes
ENSECAG00000018564	SPATS2L	0.87	0.0001	0.0221	yes
ENSECAG00000019111	CD163	0.89	0.0002	0.0308	yes

Table 3.2. Continued

ENSECAG00000020763	MGST1	0.91	0.0001	0.0221	yes
ENSECAG00000014422	OAS2	0.92	0.0001	0.0122	yes
ENSECAG00000011776	MX1	0.93	0.0001	0.0221	yes
ENSECAG00000021989	DDX58	0.94	0.0002	0.0308	yes
ENSECAG00000003474	TTLL3	0.94	0.0001	0.0122	yes
ENSECAG00000001399	SAMD9	1.03	0.0001	0.0221	yes
ENSECAG00000014218	SIGLEC1	1.04	0.0003	0.0435	yes
ENSECAG00000008274	CLEC4E	1.05	0.0001	0.0122	yes
ENSECAG00000010117	S100A9	1.05	0.0001	0.0122	yes
ENSECAG00000013762	NCR1	1.07	0.0002	0.0308	yes
ENSECAG00000023733	MMP-1	1.10	0.0001	0.0221	yes
ENSECAG00000015006	FGFR1	1.10	0.0003	0.0435	yes
ENSECAG00000007133	TPPP3	1.11	0.0001	0.0122	yes
ENSECAG00000022042	PNP	1.11	0.0001	0.0122	yes
ENSECAG00000008356	ZNF577	1.15	0.0003	0.0435	yes
ENSECAG00000012235	BAZ2B	1.15	0.0001	0.0122	yes
ENSECAG00000009742	S100A12	1.15	0.0001	0.0122	yes
ENSECAG00000021476	MMP8	1.16	0.0001	0.0122	yes
ENSECAG00000009271	S100A8	1.18	0.0001	0.0122	yes
ENSECAG00000019411	HERC6	1.20	0.0001	0.0122	yes
ENSECAG00000026820	SEPP1	1.22	0.0001	0.0122	yes
ENSECAG00000007881	IFIH1	1.24	0.0001	0.0122	yes
ENSECAG00000014645	OASL	1.29	0.0001	0.0122	yes
ENSECAG00000010153	IFIT4	1.33	0.0001	0.0122	yes
ENSECAG00000013874	SPARC	1.44	0.0002	0.0308	yes
ENSECAG00000001481	SAMD9L	1.47	0.0001	0.0122	yes
ENSECAG00000008809	OAS3	1.51	0.0001	0.0122	yes
ENSECAG00000000500	IF16	1.53	0.0001	0.0122	yes
ENSECAG00000015395	HERC5	1.54	0.0001	0.0122	yes
ENSECAG00000001324	ISG15	1.55	0.0001	0.0122	yes
ENSECAG00000008594	BTN3A1	1.59	0.0001	0.0122	yes
ENSECAG00000006913	N/A	1.90	0.0001	0.0122	yes
ENSECAG00000019949	CYP4F	1.92	0.0001	0.0122	yes
ENSECAG00000017437	MYLPF	2.38	0.0002	0.0371	yes
ENSECAG00000004349	IFIT5	2.54	0.0001	0.0122	yes
ENSECAG00000020407	MYBPC2	2.64	0.0002	0.0371	yes
ENSECAG00000010020	HBB	2.73	0.0001	0.0122	yes
ENSECAG00000023971	TNNT3	4.46	0.0002	0.0371	yes
ENSECAG00000019728	TNNC2	5.11	0.0001	0.0122	yes

Table 3.2. Continued

ENSECAG00000005487	N/A	8.08	0.0001	0.0122	yes

Results

Transcriptome-wide differential gene expression (untargeted analysis)

RNA-Seq read counts were generated for each sample at all gene loci using the ensemble gene annotation reference resulting in counts for all samples at 26,991 genes. We removed 8,212 genes for which all samples had 0 reads mapped, leaving 18,779 genes for further analysis. The first analysis performed was a comparison of the AA genotype foals and the non-AA genotyped foals. This comparison revealed 3 genes (*ANKRD2, DQB,* and *MPO;* Table 3.1) that were identified as differentially expressed with an FDR ≤ 0.05. All 3 of these genes were up-regulated in the AA genotyped foals. The second analysis of pairwise comparisons among the genotypes (*viz.,* AA vs AB, AB vs BB, and AA vs BB) did not identify any differentially expressed genes. The third analysis comparing the 2 horses that had clinical pneumonia as foals and the 10 horses that had not developed clinical signs as foals revealed no differentially expressed genes. *Cuffdiff* was then used to test for transcriptome-wide differential gene expression between the AA and non-AA genotyped samples. The *cuffdiff* analysis identified 58 differentially expressed genes with an FDR ≤ 0.05 (Table 3.2). Inspection of the results from the 2 transcriptome-wide analyses comparing the AA with the non-AA genotypes

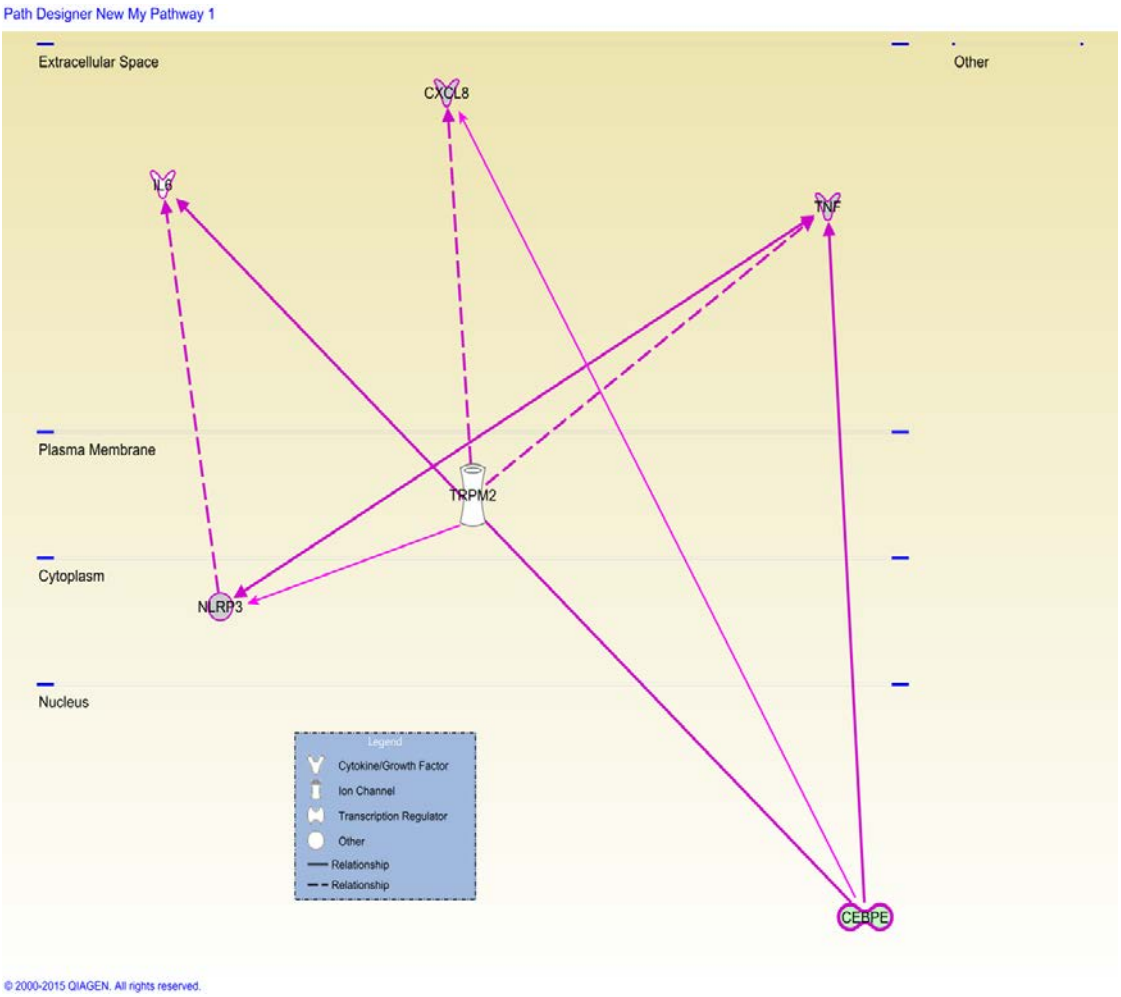


Figure 3.1. Results of the pathway analysis. *TRPM2* was not identified as differentially expressed in the analysis but was added for the purpose of identifying *TRPM2's* link with the differentially expressed genes.

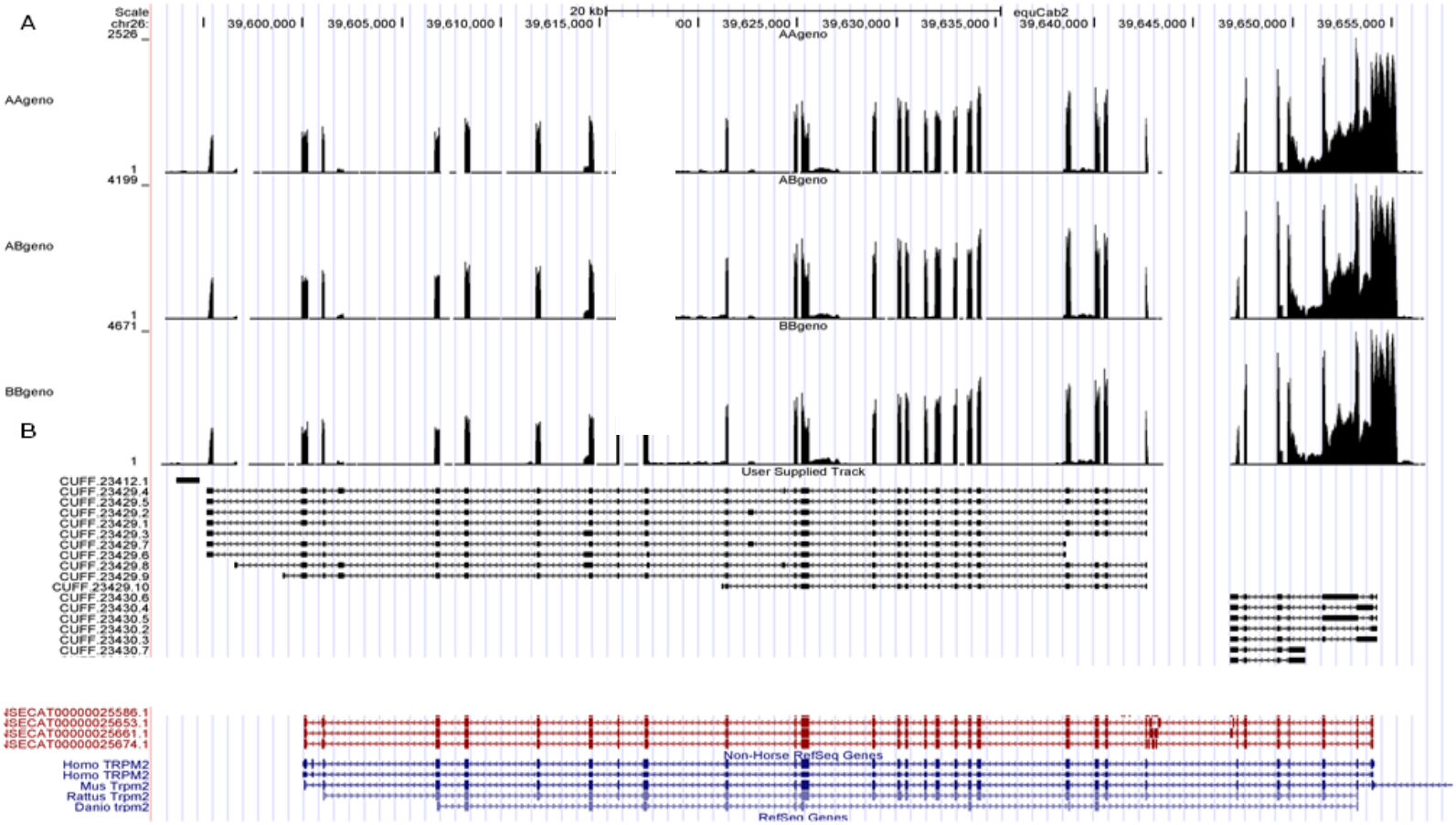


Figure 3.2. (A) Bedgraphs representing RNA-Seq coverage of *TRPM2* for each genotype. (B) Cufflinks output of *TRPM2* novel isoform overlaying the equine Ensembl predicted and RefSeq gene annotations.

using *edgeR* and *cuffdiff* revealed that no genes were concordantly identified as being differentially expressed. Expression levels of *DQB* was closest to being differentially expressed (P = 0.00075, FDR = 0.108) in the *cuffdiff* analysis, with the *ANKRD22* and *MPO* having FDRs of 0.73 and 0.63 respectively.

Looking first at the 3 genes identified as differentially expressed in the *edgeR* analysis, IPA was unable to link *ANKRD22, DQB,* and *MPO* through any pathway. Using the 58 differentially expressed genes found in the *cuffdiff* analysis, IPA was able to build a pathway which linked the differentially expressed genes (Figure 3.1). The top pathway identified included 17 of the differentially expressed genes and was related to host antimicrobial and inflammatory response as well as cell-to-cell signaling and interaction. One of the top up-regulated genes in the horses with AA genotypes compared to the non-AA genotypes was CCAAT/enhancer binding protein epsilon (*C/EBPE*) which had a 1.3-fold increase in expression over non-AA genotyped horses and was a critical component of the identified pathway structure.

Targeted analysis

Visualization of RNA-Seq coverage across the region on chromosome 26 was performed by creating bedgraphs (Figure 3.2 A) which revealed some potential novel transcripts, exons, and 3` UTRs. These findings warranted further investigation within the region of interest as novel transcripts and isoform-specific expression were not accounted for in the transcriptome-wide analyses.

To identify novel exons and transcripts we used *cufflinks* to create a new gene model representing the region around and including *TRPM2*. Transcriptome-wide, *cufflinks* assembled 34,294 transcripts and output generated was visualized in the UCSC genome browser. At the *TRPM2* locus on chromosome 26, *cufflinks* output showed the program had assembled 2 distinct genes instead of the single *TRPM2* gene currently annotated in the browser (Figure 3.2 B), that will be referred to as *TRPM2-1* and *TRMP2-2*. The currently annotated *TRPM2* gene was broken by *cufflinks* into a novel predicted transcript starting upstream of the reference annotation (*TRPM2-1)*. The remainder of the currently annotated *TRPM2* locus was incorporated by *cufflinks* into a novel independent transcript extending to the annotated 3` UTR of *TRPM2 (TRPM2-2)*. Using the 2 novel transcripts, *TRPM2-1* and *TRPM2-2*, in place of the currently annotated *TRPM2*, we re-analyzed the RNA-Seq data. The data were mapped with *HTSeq* using these 2 new coordinates for *TRPPM2* and differential expression analysis performed with *edgeR,* and no differential expression identified.

With this new gene model of *TRPM2* (*viz., TRPM2-1* and *TRPM2-2*) and using the *cuffdiff* function, we extracted only the read mapping to the new transcripts. There was significant differential expression between the AA and AB genotypes with respect to the *TRPM2-1* transcript (Table 3.3).

To investigate whether the genotypes were a marker for isoform expression differences we used the *cuffdiff* option in *Tophat* to analyze the sequencing data. Our previous analysis assigned reads to overall genes and not isoforms, whereas *cuffdiff* also allows for reads to be assigned to isoforms in order to estimate transcript abundances. When extracting only the reads for the 2 novel *TRPM2* transcripts identified with *cufflinks* we identified 10 isoforms at the *TRPM2-1* locus and 7 isoforms at the *TRPM2-2* locus. This analysis showed 1 significant difference in transcript abundances between the AA and AB genotypes at the novel *TRPM2-1* and *TRPM2-2* genes which corresponds with the above transcript differences (Table 3.3).

Table 3.3. Novel *TRPM2* transcripts pairwise comparisons.

Gene ID	Transcript ID	AA	ABBB	Log_2 FC	P value	Q value
CUFF.23429	*TRPM2-1*	90.99	118.90	0.39	0.0971	0.9995
CUFF.23430	*TRPM2-2*	168.23	214.17	0.35	0.1630	0.9995
		AA	**AB**			
CUFF.23429	*TRPM2-1*	91.81	130.55	0.51	0.0224	0.9238
CUFF.23430	*TRPM2-2*	169.73	225.94	0.41	0.0783	0.9997
		AA	**BB**			
CUFF.23429	*TRPM2-1*	90.63	110.04	0.28	0.2663	0.9991
CUFF.23430	*TRPM2-2*	167.57	205.44	0.29	0.2766	0.9991
		AB	**BB**			
CUFF.23429	*TRPM2-1*	128.83	110.27	-0.22	0.3456	0.9999
CUFF.23430	*TRPM2-2*	223.02	205.87	-0.12	0.6565	0.9999

Discussion

Biological implications

The findings reported in this paper are an additional piece of highly compelling evidence of the role of the innate immune response and the subsequent host outcome after infection by *R. equi.* As previously reported, *TRPM2* markers have been shown to

associate with clinical disease caused by *R. equi*. The TRPM2 gene was biologically plausible as a candidate gene because TRPM2 has been shown to increase tissue damage at sites of inflammation in a mouse model of ulcerative colitis [52,54]. Also, it has been demonstrated *in vitro* that a short isoform of *TRPM2* inhibits calcium influx while increasing cell viability [137]. On the basis of these findings, we considered it important to understand the expression pattern across the region of interest identified in our previous GWAS. We found evidence indicating that alternative splicing likely occurs within the *TRPM2* locus in horses resulting in multiple isoforms. Although we did not attempt to verify the functionality of these transcripts in horses, an alternative transcript has been functionally characterized in human-derived cell lines [137]; thus it is plausible that these equine splice variants might differ functionally [137]. Considering the genes identified as being differentially expressed by *cuffdiff*, 1 gene was of great interest, *viz.*, *C/EBPE*. The *C/EBP* gene and its isoforms are known to be expressed by human neutrophils and play an integral part in inducing several inflammatory cytokines [148]. C/EBP was shown to bind the promoter of interleukin 8 (*IL-8, CXCL8*) when stimulated by LPS and shows a direct link between *C/EBP* expression and the innate immune system. Another study has also highlighted the role of *TRPM2* in positively regulating *CXCL8* production leading to increased tissue damage at sites of inflammation when compared to *TRPM2*-defficient mice [54]. Another potential pathway to cell death and excessive tissue damage could be a result of the accumulation on intracellular iron (FE^{2+}) which has been shown to sensitize TRPM2 channels resulting in an upregulation of *CXCL8* production [149].

Differential gene expression analyses

The first analysis using compared the samples based on a horse's phenotypic classification when it was a foal. This was done to identify any potential changes in expression that occurred as a result of disease status while foals. This did not appear to be the case as there was no evidence that disease status as a foal had any effect on expression as an adult horse. The transcriptome-wide analysis using both *edgeR* and *cuffdiff* did not reveal any concordant results. This may be attributable to several factors. Each program uses a different method to predict differentially expressed genes and each is tailored to conduct different types of analysis. *Cuffdiff* uses a mixture of distributions to account for the uncertainty in mapping a read and the variability in read count while *edgeR* primarily focuses on the variability in read count across replicates [84,141]. The results from *edgeR* appear to be more conservative in our experiment as only 3 genes were identified as differentially expressed in comparison to *cuffdiff* which identified 58 genes. There is also evidence to suggest that *edgeR* is not always a more conservative approach; thus, the observed discrepancy might reflect some other biological or technical property present in the data but no other metrics suggest this to be the case [150]. Also, *cuffdiff* uses the accepted_hits.bam files and a gene annotation file to calculate differential expression. This removes the *HTSeq* count step and might have contributed to the differing results. In an effort to maximize the inferences drawn from our data, we elected to use the genes identified by *cuffdiff* to better leverage our data in elucidating biological pathways and processes playing a role in susceptibility to *R. equi* pneumonia.

There are several other factors to consider regarding RNA-Seq and the outcome of its analysis. RNA-Seq reflects steady state levels of RNA which encompasses the rate of transcription, rate of degradation, post-transcriptional phases, and post-translational modifications [151]. For example, a 1.3-fold increase in expression of *C/EBP* within the AA genotypes group over the non-AA genotypes is a complex finding, as it is in any case investigating differentially expressed genes. A first step in following up on the findings and better understanding this gene and pathway as it relates to *R. equi* pneumonia would be to confirm the RNA expressions levels by another method. Also, we do not know if a 1.3-fold increase in RNA correlates and indicates a 1.3 fold increase in the protein level of C/EBP in these horses. A recent study found that on average only 40% of the variability in protein levels of the cell could be explained by mRNA levels [152]. This finding shows the dynamic processes of transcription and translation cannot be easily generalized as to their functional implications.

Limitations

One limitation of this study is that it was conducted in mature horses and not in foals: gene expression as adults may not reflect the gene expression of these horses as foals when they were susceptible to *R. equi*; however, the marker genotypes will not have changed (*i.e.*, AA, AB, BB) and therefore their impact on functional transcription should remain the same. To ensure that clinical status as a foal was not impacting differential gene expression an adult, we used their respective phenotypic classifier as foals to group them for a differential expression analysis. A second limitation is that the region of

interest's involvement with and relationship to *R. equi* infection may be of a tissue-specific nature (*e.g.*, in the lung or in alveolar macrophages) and thus not captured or accurately reflected by RNA-Seq data from whole blood. A third limitation is the small sample size of the study but note that our sample size is dictated by funding. Previous studies have yielded important results using fewer horses and these presently reported results should be considered relevant to the understanding of susceptibility to *R. equi* [86,87,94]. Despite these limitations, we believe the results and contribution to knowledge of this disease outweigh these limitations.

CHAPTER IV

CONCLUSIONS

Genetic evidence regarding foal susceptibility to *R. equi* pneumonia is very clear in its common theme of implicating innate immune responses. It is certainly the case that susceptibility to *R. equi* pneumonia is controlled not by one, but by many genes within the horse and it is important to consider all findings regarding *R.equi* together when drawing conclusions. The findings presented here further strengthen that theme as concordant results were also found. The literature, which has been reviewed and presented here, agrees with the findings of this work. The first studies identifying genes associated with *R. equi* pneumonia were candidate gene studies and predominantly conducted using the Thoroughbred breed. Although the same associations were not made within the analysis of these data sets, overall themes of innate immunity genes and pathways were identified. This is an important finding as this work was conducted in Quarter horses. Generating supporting evidence across multiple breeds is important and highly compelling evidence that foal susceptibility to *R.equi* pneumonia is regulated through innate immune pathways. The importance of neutrophil function and signaling was a common theme found by these analyses.

This research also generated new hypothesis that will require further attention to identify the role of genetic variation in susceptibility to *R. equi* pneumonia. Confirmation of the isoform expression at the *TRPM2* locus as well as further evidence supporting the gene model will be required. The evidence presented in the human literature warrants further

investigation of this possibility in the horse as well. Findings that are generated by future studies will be essential steps in identifying therapeutic targets and aiding in the discovery of novel treatments for *R. equi* pneumonia.

REFERENCES

1. Prescott JF. Rhodococcus equi: an animal and human pathogen. Clin Microbiol Rev 1991;4:20-34.

2. Hurley JR, Begg AP. Failure of hyperimmune plasma to prevent pneumonia caused by Rhodococcus equi in foals. Aust Vet J 1995;72:418-420.

3. Venner M, Reinhold B, Beyerbach M, et al. Efficacy of azithromycin in preventing pulmonary abscesses in foals. Vet J 2009;179:301-303.

4. Chaffin MK, Cohen ND, Martens RJ, et al. Evaluation of the efficacy of gallium maltolate for chemoprophylaxis against pneumonia caused by Rhodococcus equi infection in foals. Am J Vet Res 2011;72:945-957.

5. Burton AJ, Giguere S, Sturgill TL, et al. Macrolide- and rifampin-resistant Rhodococcus equi on a horse breeding farm, Kentucky, USA. Emerg Infect Dis 2013;19:282-285.

6. Giguere S, Cohen ND, Chaffin MK, et al. Rhodococcus equi: clinical manifestations, virulence, and immunity. Journal of veterinary internal medicine / American College of Veterinary Internal Medicine 2011;25:1221-1230.

7. Giguere S, Hondalus MK, Yager JA, et al. Role of the 85-kilobase plasmid and plasmid-encoded virulence-associated protein A in intracellular survival and virulence of Rhodococcus equi. Infect Immun 1999;67:3548-3557.

8. Jain S, Bloom BR, Hondalus MK. Deletion of vapA encoding Virulence Associated Protein A attenuates the intracellular actinomycete Rhodococcus equi. Mol Microbiol 2003;50:115-128.

9. Cohen ND, Smith KE, Ficht TA, et al. Epidemiologic study of results of pulsed-field gel electrophoresis of isolates of Rhodococcus equi obtained from horses and horse farms. Am J Vet Res 2003;64:153-161.

10. Bolton T, Kuskie K, Halbert N, et al. Detection of strain variation in isolates of Rhodococcus equi from an affected foal using repetitive sequence-based polymerase chain reaction. J Vet Diagn Invest 2010;22:611-615.

11. Morton AC, Begg AP, Anderson GA, et al. Epidemiology of Rhodococcus equi strains on Thoroughbred horse farms. Appl Environ Microbiol 2001;67:2167-2175.

12. Cohen ND, Carter CN, Scott HM, et al. Association of soil concentrations of Rhodococcus equi and incidence of pneumonia attributable to Rhodococcus equi in foals on farms in central Kentucky. Am J Vet Res 2008;69:385-395.

13. Muscatello G, Anderson GA, Gilkerson JR, et al. Associations between the ecology of virulent Rhodococcus equi and the epidemiology of R. equi pneumonia on Australian thoroughbred farms. Appl Environ Microbiol 2006;72:6152-6160.

14. Boyd NK, Cohen ND, Lim WS, et al. Temporal changes in cytokine expression of foals during the first month of life. Vet Immunol Immunopathol 2003;92:75-85.

15. Breathnach CC, Sturgill-Wright T, Stiltner JL, et al. Foals are interferon gamma-deficient at birth. Vet Immunol Immunopathol 2006;112:199-209.

16. Stewart GR, Robertson BD, Young DB. Tuberculosis: a problem with persistence. Nat Rev Microbiol 2003;1:97-105.

17. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 2002;3:391-397.

18. Kwon JM, Goate AM. The candidate gene approach. Alcohol Res Health 2000;24:164-168.

19. Mousel MR, Harrison L, Donahue JM, et al. Rhodococcus equi and genetic susceptibility: assessing transferrin genotypes from paraffin-embedded tissues. J Vet Diagn Invest 2003;15:470-472.

20. de Jong G, van Dijk JP, van Eijk HG. The biology of transferrin. Clinica Chimica Acta 1990;190:1-46.

21. Skaar EP. The battle for iron between bacterial pathogens and their vertebrate hosts. PLoS Pathog 2010;6:e1000949.

22. Horin P, Smola J, Matiasovic J, et al. Polymorphisms in equine immune response genes and their associations with infections. Mamm Genome 2004;15:843-850.

23. Blackwell JM, Goswami T, Evans CA, et al. SLC11A1 (formerly NRAMP1) and disease resistance. Cell Microbiol 2001;3:773-784.

24. Blackwell JM, Searle S, Mohamed H, et al. Divalent cation transport and susceptibility to infectious and autoimmune disease: continuation of the Ity/Lsh/Bcg/Nramp1/Slc11a1 gene story. Immunol Lett 2003;85:197-203.

25. Vidal S, Tremblay ML, Govoni G, et al. The Ity/Lsh/Bcg locus: natural resistance to infection with intracellular parasites is abrogated by disruption of the Nramp1 gene. J Exp Med 1995;182:655-666.

26. Adams LG, Templeton JW. Genetic resistance to bacterial diseases of animals. Rev Sci Tech 1998;17:200-219.

27. Forbes JR, Gros P. Divalent-metal transport by NRAMP proteins at the interface of host-pathogen interactions. Trends Microbiol 2001;9:397-403.

28. Blackwell JM, Searle S, Goswami T, et al. Understanding the multiple functions of Nramp1. Microbes Infect 2000;2:317-321.

29. Halbert ND, Cohen ND, Slovis NM, et al. Variations in equid SLC11A1 (NRAMP1) genes and associations with Rhodococcus equi pneumonia in horses. Journal of veterinary internal medicine / American College of Veterinary Internal Medicine 2006;20:974-979.

30. Horin P, Sabakova K, Futas J, et al. Immunity-related gene single nucleotide polymorphisms associated with Rhodococcus equi infection in foals. Int J Immunogenet 2010;37:67-71.

31. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.

32. Wade CM, Giulotto E, Sigurdsson S, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 2009;326:865-867.

33. Elsik CG, Tellam RL, Worley KC, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 2009;324:522-528.

34. Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520-562.

35. Consortium TGP. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061-1073.

36. Gibbs RA, Taylor JF, Van Tassell CP, et al. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 2009;324:528-532.

37. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic acids research 2009;37:4181-4193.

38. Barsh GS, Copenhaver GP, Gibson G, et al. Guidelines for genome-wide association studies. PLoS Genet 2012;8:e1002812.

39. McCue ME, Bannasch DL, Petersen JL, et al. A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. PLoS Genet 2012;8:e1002451.

40. Finno CJ, Bannasch DL. Applied equine genetics. Equine Vet J 2014;46:538-544.

41. Go YY, Bailey E, Cook DG, et al. Genome-wide association study among four horse breeds identifies a common haplotype associated with in vitro CD3+ T cell susceptibility/resistance to equine arteritis virus infection. J Virol 2011;85:13174-13184.

42. Hauswirth R, Haase B, Blatter M, et al. Mutations in MITF and PAX3 cause "splashed white" and other white spotting phenotypes in horses. PLoS Genet 2012;8:e1002653.

43. Hill EW, McGivney BA, Gu J, et al. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. BMC Genomics 2010;11:552.

44. Lykkjen S, Dolvik NI, McCue ME, et al. Genome-wide association analysis of osteochondrosis of the tibiotarsal joint in Norwegian Standardbred trotters. Anim Genet 2010;41 Suppl 2:111-120.

45. Raudsepp T, McCue ME, Das PJ, et al. Genome-wide association study implicates testis-sperm specific FKBP6 as a susceptibility locus for impaired acrosome reaction in stallions. PLoS Genet 2012;8:e1003139.

46. Lykkjen S, Dolvik NI, McCue ME, et al. Equine developmental orthopaedic diseases--a genome-wide association study of first phalanx plantar osteochondral fragments in Standardbred trotters. Anim Genet 2013;44:766-769.

47. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet 2008;9:477-485.

48. Petersen JL, Mickelson JR, Cothran EG, et al. Genetic diversity in the modern horse illustrated from genome-wide SNP data. PloS one 2013;8:e54997.

49. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999;22:139-144.

50. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. Nature 2007;447:655-660.

51. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008;118:1590-1605.

52. McQueen CM, Doan R, Dindot SV, et al. Identification of Genomic Loci Associated with Rhodococcus equi Susceptibility in Foals. PloS one 2014;9:e98710.

53. Chaffin MK, Cohen, N.D., Blodgett, G.P., *et al.,* . Evaluation of hematologic screening methods for predicting subsequent onset of clinically apparent *Rhodococcus equi* pneumonia in foals. In: Proceeding of the 59th Annual Convention of the American Association of Equine Practitioners, Nashville 2013;267.

54. Yamamoto S, Shimizu S, Kiyonaka S, et al. TRPM2-mediated Ca2+influx induces chemokine production in monocytes that aggravates inflammatory neutrophil infiltration. Nat Med 2008;14:738-747.

55. Hastings PJ, Lupski JR, Rosenberg SM, et al. Mechanisms of change in gene copy number. Nat Rev Genet 2009;10:551-564.

56. Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet 2009;5:e1000327.

57. Lupski JR. Structural Variation in the Human Genome. New England Journal of Medicine 2007;356:1169-1171.

58. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006;7:85-97.

59. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. Trends Genet 2002;18:74-82.

60. Weterings E, van Gent DC. The mechanism of non-homologous end-joining: a synopsis of synapsis. DNA Repair (Amst) 2004;3:1425-1435.

61. Haviv-Chesner A, Kobayashi Y, Gabriel A, et al. Capture of linear fragments at a double-strand break in yeast. Nucleic acids research 2007;35:5192-5202.

62. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 2008;24:529-538.

63. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 2007;131:1235-1247.

64. Sheen CR, Jewell UR, Morris CM, et al. Double complex mutations involving F8 and FUNDC2 caused by distinct break-induced replication. Hum Mutat 2007;28:1198-1206.

65. Dupuis MC, Zhang Z, Durkin K, et al. Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. Anim Genet 2013;44:206-208.

66. Metzger J, Philipp U, Lopes MS, et al. Analysis of copy number variants by three detection algorithms and their association with body size in horses. BMC Genomics 2013;14:487.

67. Kato M, Kawaguchi T, Ishikawa S, et al. Population-genetic nature of copy number variations in the human genome. Hum Mol Genet 2010;19:761-773.

68. Mills RE, Walter K, Stewart C, et al. Mapping copy number variation by population-scale genome sequencing. Nature 2011;470:59-65.

69. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 2007;39:S16-21.

70. Doan R, Cohen ND, Sawyer J, et al. Whole-genome sequencing and genetic variant analysis of a Quarter Horse mare. BMC Genomics 2012;13:78.

71. Doan R, Cohen N, Harrington J, et al. Identification of copy number variants in horses. Genome Res 2012;22:899-907.

72. Ghosh S, Qu Z, Das PJ, et al. Copy number variation in the horse genome. PLoS Genet 2014;10:e1004712.

73. Wang W, Wang S, Hou C, et al. Genome-wide detection of copy number variations among diverse horse breeds by array CGH. PloS one 2014;9:e86860.

74. Park KD, Kim H, Hwang JY, et al. Copy number deletion has little impact on gene expression levels in racehorses. Asian-Australas J Anim Sci 2014;27:1345-1354.

75. Rosengren Pielberg G, Golovko A, Sundstrom E, et al. A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. Nat Genet 2008;40:1004-1009.

76. Olsson M, Meadows JR, Truve K, et al. A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. PLoS Genet 2011;7:e1001332.

77. Hehir-Kwa JY, Pfundt R, Veltman JA, et al. Pathogenic or not? Assessing the clinical relevance of copy number variants. Clin Genet 2013;84:415-421.

78. Krepischi AC, Pearson PL, Rosenberg C. Germline copy number variations and cancer predisposition. Future Oncol 2012;8:441-450.

79. Liu GE, Bickhart DM. Copy number variation in the cattle genome. Funct Integr Genomics 2012;12:609-624.

80. Metzker ML. Emerging technologies in DNA sequencing. Genome Res 2005;15:1767-1776.

81. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 1977;74:5463-5467.

82. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010;11:31-46.

83. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.

84. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protocols 2012;7:562-578.

85. Katz Y, Wang ET, Airoldi EM, et al. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Meth 2010;7:1009-1015.

86. Capomaccio S, Vitulo N, Verini-Supplizi A, et al. RNA sequencing of the exercise transcriptome in equine athletes. PloS one 2013;8:e83504.

87. Park KD, Park J, Ko J, et al. Whole transcriptome analyses of six thoroughbred horses before and after exercise using RNA-Seq. BMC Genomics 2012;13:473.

88. Bellone RR, Holl H, Setaluri V, et al. Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. PloS one 2013;8:e78280.

89. Das PJ, McCarthy F, Vishnoi M, et al. Stallion sperm transcriptome comprises functionally coherent coding and regulatory RNAs as revealed by microarray analysis and RNA-seq. PloS one 2013;8:e56535.

90. Iqbal K, Chitwood JL, Meyers-Brown GA, et al. RNA-seq transcriptome profiling of equine inner cell mass and trophectoderm. Biol Reprod 2014;90:61.

91. Leise BS, Watts MR, Roy S, et al. Use of laser capture microdissection for the assessment of equine lamellar basal epithelial cell signalling in the early stages of laminitis. Equine Vet J 2015;47:478-488.

92. Park W, Kim J, Kim HJ, et al. Investigation of de novo unique differentially expressed genes related to evolution in exercise response during domestication in Thoroughbred race horses. PloS one 2014;9:e91418.

93. Peffers M, Liu X, Clegg P. Transcriptomic signatures in cartilage ageing. Arthritis Res Ther 2013;15:R98.

94. Coleman SJ, Zeng Z, Wang K, et al. Structural annotation of equine protein-coding genes determined by mRNA sequencing. Anim Genet 2010;41 Suppl 2:121-130.

95. Moreton J, Malla S, Aboobaker AA, et al. Characterisation of the horse transcriptome from immunologically active tissues. PeerJ 2014;2:e382.

96. Hanash S. Disease proteomics. Nature 2003;422:226-232.

97. Chaffin MK, Cohen ND, Martens RJ. Evaluation of equine breeding farm characteristics as risk factors for development of Rhodococcus equi pneumonia in foals. J Am Vet Med Assoc 2003;222:467-475.

98. Ainsworth DM, Eicker SW, Yeagar AE, et al. Associations between physical examination, laboratory, and radiographic findings and outcome and subsequent racing performance of foals with Rhodococcus equi infection: 115 cases (1984-1992). J Am Vet Med Assoc 1998;213:510-515.

99. Ardans AA HS, Spensley MS, et al. Studies of naturally occuring and experimental *Rhodococcus equi*. Proc Am Assoc Equine Pract 1986;32:129-144.

100. Slovis NM, McCracken JL, Mundy G. How to use thoracic ultrasound to screen foals for Rhodococcus equi at affected farms. In. Lexington: Am Assoc Equine Pract; 2005:274-278.

101. Martens RJ, Martens JG, Fiske RA, et al. Rhodococcus equi foal pneumonia: protective effects of immune plasma in experimentally infected foals. Equine Vet J 1989;21:249-255.

102. Signer-Hasler H, Flury C, Haase B, et al. A genome-wide association study reveals loci influencing height and other conformation traits in horses. PloS one 2012;7:e37282.

103. Hou Y, Bickhart D, Chung H, et al. Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. Funct Integr Genomics:1-7.

104. Corbin LJ, Blott SC, Swinburne JE, et al. A genome-wide association study of osteochondritis dissecans in the Thoroughbred. Mamm Genome 2012;23:294-303.

105. Petersen JL, Mickelson JR, Rendahl AK, et al. Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLoS Genet 2013;9:e1003211.

106. Brosnahan MM, Brooks SA, Antczak DF. Equine clinical genomics: A clinician's primer. Equine Vet J 2010;42:658-670.

107. Graves KT, Henney PJ, Ennis RB. Partial deletion of the LAMA3 gene is responsible for hereditary junctional epidermolysis bullosa in the American Saddlebred Horse. Anim Genet 2009;40:35-41.

108. Lee T, Cho S, Seo KS, et al. Genetic variants and signatures of selective sweep of Hanwoo population (Korean native cattle). BMB Rep 2013;46:346-351.

109. Luo J, Yu Y, Mitra A, et al. Genome-wide copy number variant analysis in inbred chickens lines with different susceptibility to Marek's disease. G3 (Bethesda) 2013;3:217-223.

110. Hou Y, Liu GE, Bickhart DM, et al. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. Funct Integr Genomics 2012;12:81-92.

111. Nicholas TJ, Baker C, Eichler EE, et al. A high-resolution integrated map of copy number polymorphisms within and between breeds of the modern domesticated dog. BMC Genomics 2011;12:414.

112. Gokcumen O, Babb PL, Iskow RC, et al. Refinement of primate copy number variation hotspots identifies candidate genomic regions evolving under positive selection. Genome Biol 2011;12:R52.

113. Liu GE, Brown T, Hebert DA, et al. Initial analysis of copy number variations in cattle selected for resistance or susceptibility to intestinal nematodes. Mamm Genome 2011;22:111-121.

114. Liu GE, Hou Y, Zhu B, et al. Analysis of copy number variations among diverse cattle breeds. Genome Res 2010;20:693-703.

115. Nicholas TJ, Cheng Z, Ventura M, et al. The genomic architecture of segmental duplications and associated copy number variants in dogs. Genome Res 2009;19:491-499.

116. Giguère S, Cohen ND, Keith Chaffin M, et al. Diagnosis, Treatment, Control, and Prevention of Infections Caused by Rhodococcus equi in Foals. Journal of veterinary internal medicine / American College of Veterinary Internal Medicine 2011;25:1209-1220.

117. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.

118. Consotium WTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-678.

119. Aulchenko YS, Ripke S, Isaacs A, et al. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007;23:1294-1296.

120. Vazquez AI, Bates DM, Rosa GJ, et al. Technical note: an R package for fitting generalized linear mixed models in animal breeding. J Anim Sci 2010;88:497-504.

121. Price AL, Zaitlen NA, Reich D, et al. New approaches to population stratification in genome-wide association studies. Nat Rev Genet 2010;11:459-463.

122. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research 2002;30:207-210.

123. Ye S, Dhillon S, Ke X, et al. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic acids research 2001;29:E88-88.

124. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3--new capabilities and interfaces. Nucleic acids research 2012;40:e115.

125. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988;75:800-802.

126. Cohen ND, Kuskie KR, Smith JL, et al. Association of airborne concentration of virulent Rhodococcus equi with location (stall versus paddock) and month (January through June) on 30 horse breeding farms in central Kentucky. Am J Vet Res 2012;73:1603-1609.

127. Dawson TRMY, Horohov DW, Meijer WG, et al. Current understanding of the equine immune response to Rhodococcus equi. An immunological review of R. equi pneumonia. Vet Immunol Immunopathol 2010;135:1-11.

128. Heller MC, Jackson KA, Watson JL. Identification of immunologically relevant genes in mare and foal dendritic cells responding to infection by Rhodococcus equi. Vet Immunol Immunopathol 2010;136:144-150.

129. Martens RJ, Cohen ND, Jones SL, et al. Protective role of neutrophils in mice experimentally infected with Rhodococcus equi. Infect Immun 2005;73:7040-7042.

130. Chaffin MK, Cohen ND, Martens RJ, et al. Hematologic and immunophenotypic factors associated with development of Rhodococcus equi pneumonia of foals at equine breeding farms with endemic infection. Vet Immunol Immunopathol 2004;100:33-48.

131. Tateda K, Moore TA, Newstead MW, et al. Chemokine-dependent neutrophil recruitment in a murine model of Legionella pneumonia: potential role of neutrophils as immunoregulatory cells. Infect Immun 2001;69:2017-2024.

132. Lyons MJ, Yoshimura T, McMurray DN. Interleukin (IL)-8 (CXCL8) induces cytokine expression and superoxide formation by guinea pig neutrophils infected with Mycobacterium tuberculosis. Tuberculosis (Edinb) 2004;84:283-292.

133. Bordin AI, Liu M, Nerren JR, et al. Neutrophil function of neonatal foals is enhanced in vitro by CpG oligodeoxynucleotide stimulation. Vet Immunol Immunopathol 2012;145:290-297.

134. Nerren JR, Martens RJ, Payne S, et al. Age-related changes in cytokine expression by neutrophils of foals stimulated with virulent Rhodococcus equi in vitro. Vet Immunol Immunopathol 2009;127:212-219.

135. Liu M, Liu T, Bordin A, et al. Activation of foal neutrophils at different ages by CpG oligodeoxynucleotides and Rhodococcus equi. Cytokine 2009;48:280-289.

136. Ackermann M. Acute Inflamation. In: MD McGavin JZ, ed. Pathologic Basis of Veterinary Disease, 4 ed2007:120.

137. Zhang W, Chu X, Tong Q, et al. A novel TRPM2 isoform inhibits calcium influx and susceptibility to cell death. The Journal of biological chemistry 2003;278:16222-16229.

138. McQueen CM, Dindot SV, Foster MJ, et al. Genetic Susceptibility to Rhodococcus equi. Journal of veterinary internal medicine / American College of Veterinary Internal Medicine 2015.

139. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57-63.

140. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166-169.

141. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-140.

142. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic acids research 2012;40:4288-4297.

143. Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 2007;23:2881-2887.

144. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 2008;9:321-332.

145. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic acids research 2014;42:e91.

146. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841-842.

147. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res 2002;12:996-1006.

148. Cloutier A, Guindi C, Larivee P, et al. Inflammatory cytokine production by human neutrophils involves C/EBP transcription factors. Journal of immunology (Baltimore, Md : 1950) 2009;182:563-571.

149. Shimizu S, Yonezawa R, Negoro T, et al. Sensitization of HO-induced TRPM2 activation and subsequent interleukin-8 (CXCL8) production by intracellular Fe in human monocytic U937 cells. The international journal of biochemistry & cell biology 2015.

150. Nookaew I, Papini M, Pornputtapong N, et al. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. Nucleic acids research 2012;40:10084-10097.

151. Neymotin B, Athanasiadou R, Gresham D. Determination of in vivo RNA kinetics using RATE-seq. RNA (New York, NY) 2014;20:1645-1652.

152. Schwanhausser B, Busse D, Li N, et al. Global quantification of mammalian gene expression control. Nature 2011;473:337-342.

APPENDIX 2.1

Supplementary Table 2.1. CNV regions called from foals across all phenotypic groups during CNV association analysis.

Comparison 1		**Comparison 2**		**Comparison 3**	
CNV 1	**test**	**CNV2**	**test**	**CNV3**	**test**
CNVR_1261	0.019173035	**CNVR_0753**	0.01013	**CNVR_1354**	0.010975
CNVR_1936	0.026441799	**CNVR_0873**	0.012085	**CNVR_0458**	0.026079
CNVR_0645	0.030323779	**CNVR_1936**	0.032181	**CNVR_0645**	0.029019
CNVR_1354	0.036282715	**CNVR_0391**	0.032609	**CNVR_1483**	0.039921
CNVR_0873	0.054436922	**CNVR_0179**	0.03349	**CNVR_1253**	0.04016
CNVR_1483	0.0593947	**CNVR_1261**	0.033965	**CNVR_1815**	0.041391
CNVR_0753	0.065976784	**CNVR_1905**	0.048269	**CNVR_1237**	0.048842
CNVR_1457	0.066759568	**CNVR_0868**	0.061215	**CNVR_0777**	0.049978
CNVR_1166	0.071559078	**CNVR_1084**	0.062855	**CNVR_0841**	0.053025
CNVR_1815	0.079049058	**CNVR_1148**	0.064187	**CNVR_1457**	0.054011
CNVR_1074	0.084420521	**CNVR_1074**	0.087307	**CNVR_0881**	0.060429
CNVR_0868	0.102610495	**CNVR_0090**	0.088076	**CNVR_2115**	0.063301
CNVR_1237	0.106753947	**CNVR_1190**	0.088206	**CNVR_0112**	0.063932
CNVR_1122	0.112387847	**CNVR_0643**	0.091409	**CNVR_1909**	0.069793
CNVR_0391	0.117231864	**CNVR_1189**	0.09713	**CNVR_1936**	0.074018
CNVR_1293	0.121555531	**CNVR_0915**	0.10311	**CNVR_0137**	0.074887
CNVR_1905	0.122032064	**CNVR_0708**	0.108013	**CNVR_1293**	0.078871
CNVR_0773	0.130179977	**CNVR_0890**	0.10838	**CNVR_1261**	0.081324
CNVR_0458	0.130778784	**CNVR_0218**	0.110601	**CNVR_1930**	0.082416
CNVR_2115	0.131219107	**CNVR_0465**	0.110615	**CNVR_1166**	0.084361
CNVR_1097	0.135570469	**CNVR_1900**	0.110871	**CNVR_0008**	0.087838
CNVR_1253	0.136510275	**CNVR_1458**	0.111695	**CNVR_2174**	0.089134
CNVR_0008	0.139943975	**CNVR_0463**	0.123782	**CNVR_1555**	0.092809
CNVR_1418	0.15767926	**CNVR_0525**	0.126266	**CNVR_1198**	0.094162
CNVR_2221	0.160157153	**CNVR_1081**	0.128234	**CNVR_1513**	0.098585
CNVR_0643	0.188492653	**CNVR_0645**	0.128611	**CNVR_1971**	0.111661
CNVR_1678	0.193745367	**CNVR_0094**	0.129254	**CNVR_0309**	0.11343
CNVR_1190	0.194601997	**CNVR_1881**	0.135123	**CNVR_1841**	0.114328
CNVR_0890	0.195951154	**CNVR_0743**	0.135193	**CNVR_1662**	0.119684
CNVR_0915	0.204958444	**CNVR_1122**	0.135547	**CNVR_1097**	0.12376

Supp. Tbl. 2.1. Continued

CNVR_0525	0.212064567	**CNVR_0742**	0.139554	**CNVR_1274**	0.140899
CNVR_0635	0.225165786	**CNVR_1764**	0.139793	**CNVR_1826**	0.149904
CNVR_1915	0.230353666	**CNVR_0939**	0.14053	**CNVR_0269**	0.15104
CNVR_1900	0.237406365	**CNVR_1786**	0.146216	**CNVR_0322**	0.153913
CNVR_0881	0.237700852	**CNVR_0664**	0.1475	**CNVR_2234**	0.15885
CNVR_0076	0.239895885	**CNVR_0949**	0.149696	**CNVR_2107**	0.161239
CNVR_1473	0.255928454	**CNVR_1027**	0.153132	**CNVR_2237**	0.162794
CNVR_1909	0.262603785	**CNVR_0056**	0.156166	**CNVR_0088**	0.163527
CNVR_0708	0.264237205	**CNVR_0773**	0.158931	**CNVR_0076**	0.164981
CNVR_1874	0.270338275	**CNVR_0212**	0.160537	**CNVR_1328**	0.171146
CNVR_0237	0.276379102	**CNVR_1721**	0.165236	**CNVR_1383**	0.173899
CNVR_1148	0.278418665	**CNVR_1345**	0.168864	**CNVR_1268**	0.177435
CNVR_1455	0.287915853	**CNVR_2051**	0.176407	**CNVR_1260**	0.178015
CNVR_0842	0.294342951	**CNVR_2234**	0.176412	**CNVR_1074**	0.178281
CNVR_0664	0.296260331	**CNVR_1166**	0.179104	**CNVR_0154**	0.182245
CNVR_0606	0.300079787	**CNVR_1483**	0.181291	**CNVR_0409**	0.184381
CNVR_0593	0.300656968	**CNVR_1095**	0.181427	**CNVR_2011**	0.184488
CNVR_1474	0.301525745	**CNVR_1457**	0.182906	**CNVR_1636**	0.187221
CNVR_0094	0.306255913	**CNVR_1355**	0.185885	**CNVR_1084**	0.187467
CNVR_1351	0.306754219	**CNVR_0295**	0.189518	**CNVR_0006**	0.192393
CNVR_0939	0.307035028	**CNVR_0993**	0.192112	**CNVR_1859**	0.199032
CNVR_1841	0.308594592	**CNVR_1700**	0.192922	**CNVR_0217**	0.199663
CNVR_2127	0.308704165	**CNVR_0128**	0.194288	**CNVR_2145**	0.200479
CNVR_0847	0.312907485	**CNVR_1787**	0.196726	**CNVR_0785**	0.200527
CNVR_2149	0.317208702	**CNVR_1870**	0.196915	**CNVR_1956**	0.200651
CNVR_0137	0.320305357	**CNVR_0735**	0.199355	**CNVR_2059**	0.213569
CNVR_1813	0.326881751	**CNVR_1066**	0.200972	**CNVR_0814**	0.219067
CNVR_1833	0.33054393	**CNVR_0321**	0.201898	**CNVR_2083**	0.222987
CNVR_1002	0.335267606	**CNVR_0816**	0.202033	**CNVR_1263**	0.230647
CNVR_1555	0.339863597	**CNVR_1214**	0.202204	**CNVR_2118**	0.238848
CNVR_1278	0.341389981	**CNVR_2243**	0.203174	**CNVR_1122**	0.244064
CNVR_0993	0.34256069	**CNVR_2152**	0.203279	**CNVR_0847**	0.247441
CNVR_2107	0.3430277	**CNVR_1874**	0.205366	**CNVR_0169**	0.249155
CNVR_1027	0.346081794	**CNVR_1085**	0.207509	**CNVR_1678**	0.252626
CNVR_0169	0.347238742	**CNVR_1002**	0.21149	**CNVR_2232**	0.25493
CNVR_1458	0.352084477	**CNVR_1418**	0.214819	**CNVR_0250**	0.255496
CNVR_1965	0.35355713	**CNVR_1678**	0.21793	**CNVR_0465**	0.256833
CNVR_0218	0.355955648	**CNVR_1726**	0.219095	**CNVR_2221**	0.261694
CNVR_1342	0.356569729	**CNVR_0412**	0.223336	**CNVR_1187**	0.26419

Supp. Tbl. 2.1. Continued

CNVR_1066	0.356592914	**CNVR_2221**	0.225817	**CNVR_2069**	0.266439
CNVR_1513	0.356685244	**CNVR_0635**	0.22785	**CNVR_2149**	0.273412
CNVR_1198	0.356924669	**CNVR_2089**	0.228299	**CNVR_1255**	0.276083
CNVR_1636	0.357772891	**CNVR_1702**	0.229675	**CNVR_0050**	0.279331
CNVR_0322	0.358083053	**CNVR_0632**	0.229686	**CNVR_2251**	0.281156
CNVR_1255	0.367908607	**CNVR_1407**	0.22993	**CNVR_1486**	0.28153
CNVR_1894	0.372358443	**CNVR_1859**	0.230183	**CNVR_1455**	0.283633
CNVR_2145	0.373049752	**CNVR_2054**	0.231243	**CNVR_2165**	0.284531
CNVR_0777	0.373239288	**CNVR_0794**	0.2333	**CNVR_1418**	0.284891
CNVR_1268	0.374495457	**CNVR_1915**	0.233724	**CNVR_0639**	0.287119
CNVR_1260	0.375543669	**CNVR_0462**	0.233882	**CNVR_0162**	0.287996
CNVR_0154	0.378018236	**CNVR_0536**	0.234475	**CNVR_0842**	0.28943
CNVR_1061	0.380491293	**CNVR_1278**	0.235895	**CNVR_0773**	0.289497
CNVR_0162	0.381586642	**CNVR_0894**	0.236604	**CNVR_2072**	0.290513
CNVR_0248	0.384163351	**CNVR_1246**	0.238735	**CNVR_1764**	0.295355
CNVR_0417	0.385116228	**CNVR_0785**	0.239357	**CNVR_2127**	0.296268
CNVR_2118	0.386096841	**CNVR_0558**	0.239998	**CNVR_1061**	0.298337
CNVR_0949	0.386236353	**CNVR_0637**	0.241343	**CNVR_0160**	0.299678
CNVR_1446	0.386628181	**CNVR_0738**	0.24247	**CNVR_2024**	0.301934
CNVR_0295	0.392734284	**CNVR_2165**	0.24604	**CNVR_1013**	0.304986
CNVR_0841	0.393175609	**CNVR_1473**	0.246956	**CNVR_1624**	0.306462
CNVR_0056	0.393370481	**CNVR_1317**	0.256478	**CNVR_1277**	0.31006
CNVR_2181	0.395955599	**CNVR_1077**	0.257444	**CNVR_1216**	0.312502
CNVR_1623	0.397998194	**CNVR_1144**	0.258629	**CNVR_0251**	0.319321
CNVR_0463	0.398650794	**CNVR_1640**	0.258957	**CNVR_0689**	0.326737
CNVR_1700	0.400488587	**CNVR_1838**	0.264326	**CNVR_0482**	0.328061
CNVR_1826	0.402151008	**CNVR_1304**	0.265933	**CNVR_2213**	0.329374
CNVR_1355	0.406264293	**CNVR_0722**	0.266088	**CNVR_2246**	0.329672
CNVR_1752	0.408495979	**CNVR_1097**	0.26761	**CNVR_1652**	0.335554
CNVR_1400	0.412139219	**CNVR_0606**	0.267653	**CNVR_0610**	0.335554
CNVR_0894	0.419382529	**CNVR_1138**	0.270273	**CNVR_0052**	0.340177
CNVR_0128	0.423224248	**CNVR_2128**	0.272155	**CNVR_0237**	0.341367
CNVR_2015	0.430757448	**CNVR_1107**	0.272545	**CNVR_1308**	0.342132
CNVR_1956	0.436675211	**CNVR_1813**	0.272887	**CNVR_0593**	0.343487
CNVR_0462	0.437983353	**CNVR_1370**	0.2766	**CNVR_1577**	0.34689
CNVR_1246	0.440989021	**CNVR_1188**	0.277599	**CNVR_0873**	0.346904
CNVR_1039	0.442068868	**CNVR_1469**	0.277633	**CNVR_2181**	0.349781
CNVR_0558	0.448196739	**CNVR_1023**	0.278379	**CNVR_0179**	0.350252
CNVR_1693	0.452186739	**CNVR_0331**	0.27927	**CNVR_1272**	0.35201

Supp. Tbl. 2.1. Continued

CNVR_1680	0.455263752	**CNVR_2015**	0.281794	**CNVR_1085**	0.353449
CNVR_1662	0.458555504	**CNVR_2009**	0.286619	**CNVR_1693**	0.356704
CNVR_1786	0.459389915	**CNVR_1926**	0.28678	**CNVR_1642**	0.357768
CNVR_0742	0.460382863	**CNVR_1474**	0.286833	**CNVR_2175**	0.359971
CNVR_1214	0.462444812	**CNVR_1015**	0.28734	**CNVR_1825**	0.367001
CNVR_0217	0.467192905	**CNVR_0063**	0.291507	**CNVR_2262**	0.370763
CNVR_0876	0.467295248	**CNVR_0052**	0.291649	**CNVR_1167**	0.373317
CNVR_1721	0.468517458	**CNVR_1885**	0.292599	**CNVR_2030**	0.373893
CNVR_1725	0.468737315	**CNVR_0084**	0.292785	**CNVR_1891**	0.37424
CNVR_0743	0.470975233	**CNVR_0493**	0.293246	**CNVR_0456**	0.377432
CNVR_1971	0.474121677	**CNVR_0245**	0.293608	**CNVR_1351**	0.385905
CNVR_0589	0.481494511	**CNVR_1446**	0.295415	**CNVR_0849**	0.387034
CNVR_1081	0.485372036	**CNVR_1634**	0.29689	**CNVR_2239**	0.387278
CNVR_1759	0.488655354	**CNVR_1930**	0.301234	**CNVR_2200**	0.388868
CNVR_1138	0.489468285	**CNVR_2157**	0.303455	**CNVR_1039**	0.390115
CNVR_0212	0.491676969	**CNVR_2060**	0.304606	**CNVR_0216**	0.391
CNVR_0816	0.492042257	**CNVR_1830**	0.308676	**CNVR_1189**	0.396292
CNVR_1870	0.49275206	**CNVR_1532**	0.308718	**CNVR_2148**	0.397451
CNVR_1345	0.492964883	**CNVR_1785**	0.312669	**CNVR_1915**	0.39843
CNVR_0493	0.493153277	**CNVR_1185**	0.313966	**CNVR_1879**	0.398643
CNVR_0179	0.493267691	**CNVR_1570**	0.315686	**CNVR_1914**	0.404673
CNVR_1532	0.495719888	**CNVR_0814**	0.317145	**CNVR_0895**	0.410258
CNVR_0090	0.497165008	**CNVR_1815**	0.317497	**CNVR_0875**	0.410296
CNVR_0905	0.499536659	**CNVR_1328**	0.32428	**CNVR_1058**	0.410509
CNVR_1881	0.502510476	**CNVR_0281**	0.324424	**CNVR_2060**	0.414612
CNVR_0245	0.506490746	**CNVR_1383**	0.324587	**CNVR_0660**	0.420235
CNVR_0106	0.515021679	**CNVR_0905**	0.329067	**CNVR_0432**	0.420244
CNVR_1185	0.515283499	**CNVR_1256**	0.329664	**CNVR_0876**	0.42095
CNVR_0084	0.515403738	**CNVR_1988**	0.330455	**CNVR_1894**	0.421573
CNVR_0269	0.515973621	**CNVR_1624**	0.333282	**CNVR_0632**	0.424844
CNVR_0010	0.517166067	**CNVR_1400**	0.333663	**CNVR_1965**	0.427686
CNVR_0309	0.521057179	**CNVR_0417**	0.337177	**CNVR_0570**	0.432284
CNVR_1407	0.52288646	**CNVR_0885**	0.337254	**CNVR_1833**	0.432463
CNVR_1921	0.523378684	**CNVR_1921**	0.337919	**CNVR_0395**	0.434913
CNVR_0875	0.523696262	**CNVR_0554**	0.338002	**CNVR_0940**	0.435843
CNVR_0216	0.525320715	**CNVR_2098**	0.338084	**CNVR_2038**	0.435921
CNVR_1633	0.528643404	**CNVR_2241**	0.338883	**CNVR_0149**	0.437174
CNVR_0734	0.532066367	**CNVR_1623**	0.34014	**CNVR_2089**	0.440611
CNVR_0689	0.53912549	**CNVR_1715**	0.340467	**CNVR_1473**	0.443917

Supp. Tbl. 2.1**.** Continued

CNVR_2152	0.542562558	**CNVR_1833**	0.340595	**CNVR_1983**	0.447395
CNVR_0063	0.54337445	**CNVR_1611**	0.34109	**CNVR_1359**	0.448745
CNVR_0149	0.543929363	**CNVR_1342**	0.341118	**CNVR_1107**	0.448907
CNVR_0637	0.544554965	**CNVR_1759**	0.341574	**CNVR_0635**	0.45229
CNVR_1988	0.545450591	**CNVR_0547**	0.341976	**CNVR_0868**	0.452618
CNVR_1274	0.546548081	**CNVR_2194**	0.342179	**CNVR_1832**	0.452837
CNVR_1095	0.546646775	**CNVR_0237**	0.342785	**CNVR_1752**	0.456635
CNVR_2054	0.549154429	**CNVR_1222**	0.343497	**CNVR_1787**	0.459803
CNVR_1077	0.550851809	**CNVR_0010**	0.344484	**CNVR_0571**	0.463383
CNVR_1370	0.552403756	**CNVR_0620**	0.345165	**CNVR_1707**	0.476333
CNVR_0735	0.552852954	**CNVR_1351**	0.346592	**CNVR_1108**	0.479237
CNVR_1047	0.560036206	**CNVR_0746**	0.350347	**CNVR_1633**	0.486882
CNVR_1187	0.563847152	**CNVR_0409**	0.354334	**CNVR_1533**	0.487726
CNVR_2102	0.566307761	**CNVR_0965**	0.35586	**CNVR_0470**	0.489644
CNVR_0112	0.566973515	**CNVR_1725**	0.358387	**CNVR_1342**	0.489856
CNVR_0321	0.574370336	**CNVR_2264**	0.358461	**CNVR_0919**	0.492929
CNVR_1188	0.57639788	**CNVR_2090**	0.358646	**CNVR_1119**	0.494754
CNVR_1726	0.580114693	**CNVR_0660**	0.361515	**CNVR_0224**	0.496761
CNVR_2174	0.582134018	**CNVR_0610**	0.362367	**CNVR_0106**	0.497812
CNVR_0536	0.586055988	**CNVR_1293**	0.363347	**CNVR_1222**	0.499582
CNVR_1838	0.590667282	**CNVR_0593**	0.363659	**CNVR_1304**	0.503541
CNVR_1990	0.59183767	**CNVR_1829**	0.364313	**CNVR_0331**	0.504547
CNVR_1930	0.591920252	**CNVR_0663**	0.366271	**CNVR_1474**	0.507492
CNVR_1640	0.594874442	**CNVR_1233**	0.368664	**CNVR_1905**	0.507823
CNVR_2051	0.595796923	**CNVR_0251**	0.368894	**CNVR_0320**	0.510656
CNVR_0885	0.597868322	**CNVR_0571**	0.369361	**CNVR_0915**	0.5109
CNVR_1820	0.59940217	**CNVR_0248**	0.369512	**CNVR_1647**	0.511797
CNVR_1013	0.602877082	**CNVR_1954**	0.371435	**CNVR_1317**	0.512858
CNVR_1829	0.603593319	**CNVR_0125**	0.372098	**CNVR_1935**	0.515449
CNVR_1203	0.606896359	**CNVR_0956**	0.372415	**CNVR_0606**	0.51688
CNVR_0232	0.607271064	**CNVR_1533**	0.379195	**CNVR_1444**	0.517524
CNVR_0006	0.608303438	**CNVR_0008**	0.383717	**CNVR_0502**	0.518799
CNVR_1647	0.611329121	**CNVR_1990**	0.384717	**CNVR_1144**	0.521745
CNVR_1233	0.611657908	**CNVR_0968**	0.38813	**CNVR_0090**	0.52249
CNVR_1189	0.612282582	**CNVR_1972**	0.389229	**CNVR_1203**	0.52422
CNVR_1469	0.612987289	**CNVR_0585**	0.389352	**CNVR_1972**	0.525088
CNVR_2210	0.613191851	**CNVR_1680**	0.391567	**CNVR_2090**	0.525193
CNVR_0956	0.61596022	**CNVR_1690**	0.394229	**CNVR_1340**	0.537771
CNVR_0794	0.616538033	**CNVR_2174**	0.395206	**CNVR_2245**	0.540141

Supp. Tbl. 2.1. Continued

CNVR_1570	0.619367966	**CNVR_0385**	0.405193	**CNVR_2103**	0.542406
CNVR_0738	0.62107015	**CNVR_2071**	0.406967	**CNVR_2202**	0.542557
CNVR_0250	0.622889942	**CNVR_0748**	0.408374	**CNVR_1821**	0.542625
CNVR_2059	0.624850971	**CNVR_0821**	0.408958	**CNVR_0075**	0.544846
CNVR_1702	0.627319642	**CNVR_2072**	0.408982	**CNVR_1813**	0.546327
CNVR_2237	0.627567086	**CNVR_1047**	0.411082	**CNVR_0643**	0.547823
CNVR_1023	0.630124004	**CNVR_0895**	0.41109	**CNVR_2009**	0.548626
CNVR_0852	0.630878527	**CNVR_0470**	0.411409	**CNVR_0525**	0.553185
CNVR_2243	0.632486122	**CNVR_1354**	0.412799	**CNVR_1874**	0.555976
CNVR_0908	0.635212908	**CNVR_1161**	0.412821	**CNVR_1634**	0.559665
CNVR_0125	0.63527982	**CNVR_1263**	0.413166	**CNVR_1978**	0.561392
CNVR_2232	0.637466332	**CNVR_1167**	0.415639	**CNVR_0248**	0.564226
CNVR_1088	0.638944176	**CNVR_1978**	0.417512	**CNVR_0616**	0.564886
CNVR_0221	0.645103991	**CNVR_1652**	0.418163	**CNVR_2098**	0.567012
CNVR_0034	0.646259493	**CNVR_0940**	0.424137	**CNVR_0412**	0.567729
CNVR_2200	0.647394944	**CNVR_1340**	0.426668	**CNVR_1164**	0.567984
CNVR_2024	0.647723764	**CNVR_1455**	0.426749	**CNVR_1836**	0.569905
CNVR_0570	0.648628347	**CNVR_2039**	0.42774	**CNVR_0189**	0.573111
CNVR_0651	0.654081531	**CNVR_0194**	0.429715	**CNVR_0281**	0.57434
CNVR_1272	0.658196394	**CNVR_0421**	0.43189	**CNVR_0890**	0.574972
CNVR_0200	0.66107243	**CNVR_2102**	0.433798	**CNVR_0145**	0.57653
CNVR_1086	0.664093396	**CNVR_0320**	0.437989	**CNVR_0722**	0.576747
CNVR_2148	0.667319808	**CNVR_0112**	0.438181	**CNVR_1611**	0.579142
CNVR_1926	0.669018095	**CNVR_0734**	0.439008	**CNVR_2071**	0.58118
CNVR_0965	0.672905051	**CNVR_0842**	0.439343	**CNVR_1783**	0.582476
CNVR_1368	0.678221034	**CNVR_0919**	0.443136	**CNVR_0107**	0.582519
CNVR_1891	0.679591513	**CNVR_1308**	0.444706	**CNVR_1885**	0.582701
CNVR_1764	0.680452056	**CNVR_0764**	0.455049	**CNVR_1161**	0.585511
CNVR_1715	0.680456207	**CNVR_1965**	0.457717	**CNVR_2194**	0.586724
CNVR_0145	0.681782642	**CNVR_0107**	0.457998	**CNVR_2039**	0.588615
CNVR_1830	0.683626035	**CNVR_1914**	0.45813	**CNVR_0385**	0.594324
CNVR_1690	0.683900491	**CNVR_0364**	0.458269	**CNVR_2051**	0.604225
CNVR_1256	0.686665985	**CNVR_1820**	0.461279	**CNVR_0620**	0.605812
CNVR_1787	0.689945643	**CNVR_1758**	0.463878	**CNVR_0554**	0.610336
CNVR_2128	0.6900729	**CNVR_0073**	0.464786	**CNVR_0547**	0.612881
CNVR_1015	0.692257865	**CNVR_1237**	0.465354	**CNVR_1081**	0.617642
CNVR_0722	0.692819991	**CNVR_1983**	0.4655	**CNVR_1881**	0.618619
CNVR_2213	0.693279729	**CNVR_2127**	0.466021	**CNVR_1954**	0.620539
CNVR_2241	0.702703417	**CNVR_1935**	0.467551	**CNVR_1307**	0.623591

Supp. Tbl. 2.1. Continued

CNVR_1090	0.703726175	**CNVR_0088**	0.469512	**CNVR_0746**	0.625403
CNVR_1785	0.70377787	**CNVR_2011**	0.471537	**CNVR_1278**	0.626328
CNVR_0830	0.708114532	**CNVR_0589**	0.47521	**CNVR_2128**	0.628047
CNVR_0465	0.709176128	**CNVR_1090**	0.475641	**CNVR_1190**	0.629977
CNVR_0050	0.710384752	**CNVR_1752**	0.47929	**CNVR_0218**	0.630354
CNVR_0412	0.711518528	**CNVR_2115**	0.479351	**CNVR_1446**	0.631367
CNVR_0512	0.712423247	**CNVR_0177**	0.483933	**CNVR_2124**	0.631534
CNVR_0088	0.717514252	**CNVR_0512**	0.484098	**CNVR_0118**	0.633582
CNVR_1821	0.720149061	**CNVR_2030**	0.485205	**CNVR_2264**	0.636648
CNVR_2089	0.720664021	**CNVR_0777**	0.48825	**CNVR_0498**	0.639549
CNVR_1502	0.725717078	**CNVR_1486**	0.492776	**CNVR_1400**	0.641061
CNVR_1108	0.726162433	**CNVR_0076**	0.493369	**CNVR_0738**	0.64379
CNVR_2175	0.727457672	**CNVR_1894**	0.50116	**CNVR_0221**	0.644139
CNVR_1313	0.728959652	**CNVR_1088**	0.501313	**CNVR_0589**	0.647401
CNVR_1599	0.732251546	**CNVR_1374**	0.507223	**CNVR_0200**	0.647748
CNVR_1670	0.734643008	**CNVR_1577**	0.508919	**CNVR_1730**	0.648753
CNVR_1295	0.740467945	**CNVR_0616**	0.509964	**CNVR_2243**	0.650076
CNVR_2103	0.740470386	**CNVR_1502**	0.510507	**CNVR_1121**	0.654754
CNVR_0219	0.740588861	**CNVR_2119**	0.512446	**CNVR_1623**	0.656464
CNVR_1634	0.741710877	**CNVR_1121**	0.51377	**CNVR_2157**	0.6571
CNVR_0968	0.742092407	**CNVR_0034**	0.519698	**CNVR_0417**	0.665303
CNVR_1885	0.74264059	**CNVR_1730**	0.519963	**CNVR_0761**	0.665862
CNVR_2157	0.743628137	**CNVR_2237**	0.522139	**CNVR_0261**	0.666494
CNVR_1317	0.743729161	**CNVR_2149**	0.526759	**CNVR_0247**	0.670066
CNVR_0096	0.74443516	**CNVR_0502**	0.530001	**CNVR_0585**	0.671027
CNVR_1304	0.745468538	**CNVR_2251**	0.535253	**CNVR_1854**	0.671485
CNVR_0160	0.748673658	**CNVR_0224**	0.535458	**CNVR_2143**	0.671706
CNVR_0554	0.753315241	**CNVR_1436**	0.536501	**CNVR_1015**	0.672078
CNVR_0911	0.753771217	**CNVR_0232**	0.54003	**CNVR_0019**	0.674882
CNVR_0663	0.755824136	**CNVR_0456**	0.542542	**CNVR_0748**	0.676858
CNVR_2264	0.755924637	**CNVR_2069**	0.542648	**CNVR_1670**	0.677716
CNVR_1568	0.756685645	**CNVR_1030**	0.545439	**CNVR_1614**	0.679536
CNVR_0177	0.759656853	**CNVR_1444**	0.54555	**CNVR_2241**	0.680388
CNVR_2083	0.759738011	**CNVR_1216**	0.546338	**CNVR_1926**	0.68398
CNVR_1156	0.762633159	**CNVR_1599**	0.549044	**CNVR_1680**	0.685373
CNVR_0547	0.763997678	**CNVR_1736**	0.550393	**CNVR_0663**	0.68568
CNVR_2011	0.768523777	**CNVR_2210**	0.550905	**CNVR_1785**	0.687382
CNVR_0281	0.768773981	**CNVR_2083**	0.555595	**CNVR_0165**	0.687468
CNVR_0421	0.769677558	**CNVR_1836**	0.558651	**CNVR_0391**	0.689027

Supp. Tbl. 2.1. Continued

CNVR_1263	0.769703931	**CNVR_2124**	0.55918	**CNVR_0821**	0.690771
CNVR_1144	0.771909705	**CNVR_1277**	0.559302	**CNVR_0364**	0.697859
CNVR_0585	0.773739452	**CNVR_0096**	0.571865	**CNVR_0194**	0.698594
CNVR_2239	0.774280573	**CNVR_0482**	0.574268	**CNVR_0734**	0.703274
CNVR_1328	0.775490316	**CNVR_0639**	0.574839	**CNVR_1313**	0.706645
CNVR_1832	0.776302994	**CNVR_0847**	0.576804	**CNVR_0735**	0.707859
CNVR_1954	0.777671262	**CNVR_1039**	0.57681	**CNVR_1023**	0.710084
CNVR_1825	0.780678122	**CNVR_0123**	0.582182	**CNVR_0968**	0.710716
CNVR_0331	0.781936404	**CNVR_0852**	0.587895	**CNVR_1374**	0.710992
CNVR_0764	0.785159295	**CNVR_1642**	0.58939	**CNVR_0421**	0.71383
CNVR_0748	0.785689097	**CNVR_1061**	0.59097	**CNVR_1726**	0.716561
CNVR_2194	0.786341428	**CNVR_1879**	0.591476	**CNVR_0212**	0.719077
CNVR_1107	0.786440929	**CNVR_0651**	0.596852	**CNVR_1130**	0.729108
CNVR_0194	0.789486969	**CNVR_2246**	0.598963	**CNVR_0908**	0.729848
CNVR_1611	0.791505771	**CNVR_2181**	0.601334	**CNVR_1786**	0.7345
CNVR_2069	0.792605485	**CNVR_1253**	0.605718	**CNVR_1256**	0.737128
CNVR_2262	0.792949313	**CNVR_2202**	0.605898	**CNVR_0536**	0.74188
CNVR_2098	0.800839219	**CNVR_0395**	0.611426	**CNVR_2152**	0.745636
CNVR_2245	0.801535546	**CNVR_0164**	0.611436	**CNVR_0742**	0.746492
CNVR_1030	0.803568577	**CNVR_1974**	0.612761	**CNVR_1568**	0.747303
CNVR_0821	0.803595921	**CNVR_1086**	0.619782	**CNVR_1095**	0.749169
CNVR_1758	0.803728576	**CNVR_0432**	0.631866	**CNVR_0911**	0.749826
CNVR_2009	0.807544903	**CNVR_1255**	0.633369	**CNVR_1002**	0.749883
CNVR_1084	0.807902202	**CNVR_1058**	0.634475	**CNVR_0321**	0.750669
CNVR_0620	0.807928187	**CNVR_0106**	0.638502	**CNVR_1715**	0.750715
CNVR_2038	0.811631272	**CNVR_2059**	0.641257	**CNVR_1592**	0.751062
CNVR_0073	0.811658531	**CNVR_1368**	0.644378	**CNVR_0073**	0.751348
CNVR_0632	0.81211763	**CNVR_1707**	0.646114	**CNVR_1830**	0.751824
CNVR_0247	0.812733136	**CNVR_0876**	0.648167	**CNVR_1758**	0.754617
CNVR_1374	0.813280675	**CNVR_0849**	0.652087	**CNVR_0018**	0.754725
CNVR_0075	0.813750246	**CNVR_0908**	0.654305	**CNVR_0764**	0.757955
CNVR_1436	0.820908094	**CNVR_0006**	0.654853	**CNVR_1974**	0.758772
CNVR_0123	0.822410239	**CNVR_0841**	0.655251	**CNVR_2119**	0.75992
CNVR_1307	0.823115212	**CNVR_0162**	0.655554	**CNVR_0664**	0.76058
CNVR_2246	0.82358608	**CNVR_0169**	0.655608	**CNVR_0219**	0.761777
CNVR_2060	0.825295236	**CNVR_0830**	0.656211	**CNVR_1736**	0.766852
CNVR_0482	0.828303296	**CNVR_0223**	0.657222	**CNVR_0794**	0.773283
CNVR_0165	0.829246198	**CNVR_0160**	0.657222	**CNVR_1368**	0.773898
CNVR_1972	0.830200343	**CNVR_0498**	0.658321	**CNVR_1436**	0.782388

Supp. Tbl. 2.1. Continued

CNVR_0849	0.830670313	**CNVR_1119**	0.659159	**CNVR_0753**	0.782575
CNVR_1216	0.830810695	**CNVR_1295**	0.660104	**CNVR_2054**	0.785555
CNVR_0385	0.831753372	**CNVR_2239**	0.66423	**CNVR_0232**	0.785715
CNVR_0409	0.833012392	**CNVR_0189**	0.664232	**CNVR_1721**	0.786255
CNVR_1222	0.833746396	**CNVR_0309**	0.666138	**CNVR_1640**	0.788451
CNVR_2119	0.835211044	**CNVR_0458**	0.666777	**CNVR_0965**	0.788659
CNVR_1788	0.836523964	**CNVR_0261**	0.670203	**CNVR_1725**	0.789488
CNVR_2090	0.838188376	**CNVR_2175**	0.670221	**CNVR_1494**	0.793054
CNVR_0164	0.838951718	**CNVR_1156**	0.67066	**CNVR_0184**	0.79447
CNVR_0761	0.839611968	**CNVR_1582**	0.678136	**CNVR_0993**	0.795518
CNVR_1315	0.839836322	**CNVR_1825**	0.684651	**CNVR_0852**	0.795803
CNVR_1161	0.841491256	**CNVR_1359**	0.685699	**CNVR_1759**	0.797956
CNVR_0548	0.843256488	**CNVR_1693**	0.696551	**CNVR_1066**	0.799407
CNVR_2014	0.84504204	**CNVR_1164**	0.699012	**CNVR_0223**	0.800925
CNVR_2039	0.845746401	**CNVR_1783**	0.699531	**CNVR_1407**	0.804828
CNVR_0364	0.847597677	**CNVR_1315**	0.70032	**CNVR_2015**	0.807381
CNVR_2251	0.847667486	**CNVR_1274**	0.704294	**CNVR_1469**	0.807715
CNVR_1978	0.849244871	**CNVR_2262**	0.70472	**CNVR_1086**	0.809959
CNVR_0746	0.854128102	**CNVR_1633**	0.712672	**CNVR_1656**	0.814001
CNVR_2072	0.861064022	**CNVR_0118**	0.718098	**CNVR_1345**	0.824746
CNVR_1642	0.862326746	**CNVR_0881**	0.718891	**CNVR_0743**	0.828016
CNVR_0184	0.862821977	**CNVR_1854**	0.721223	**CNVR_0939**	0.828538
CNVR_1736	0.863236829	**CNVR_1614**	0.726644	**CNVR_0651**	0.831605
CNVR_0019	0.867188778	**CNVR_1343**	0.72893	**CNVR_0463**	0.834344
CNVR_0814	0.868601023	**CNVR_0154**	0.734868	**CNVR_0830**	0.837126
CNVR_2165	0.869228407	**CNVR_0200**	0.737917	**CNVR_0123**	0.841104
CNVR_0432	0.870978621	**CNVR_2038**	0.738384	**CNVR_1870**	0.84119
CNVR_1359	0.871008784	**CNVR_1832**	0.738834	**CNVR_1570**	0.842856
CNVR_1486	0.871584689	**CNVR_1909**	0.744698	**CNVR_1702**	0.843347
CNVR_1085	0.872417313	**CNVR_2118**	0.74577	**CNVR_2210**	0.844328
CNVR_2234	0.874931207	**CNVR_0221**	0.749298	**CNVR_1395**	0.849025
CNVR_0018	0.879382934	**CNVR_0050**	0.749406	**CNVR_1690**	0.853128
CNVR_0660	0.879390182	**CNVR_1788**	0.753043	**CNVR_2102**	0.854766
CNVR_2071	0.88026757	**CNVR_0250**	0.755217	**CNVR_0177**	0.855424
CNVR_1340	0.88369463	**CNVR_1494**	0.758818	**CNVR_2047**	0.855548
CNVR_0456	0.88621487	**CNVR_1260**	0.765639	**CNVR_1838**	0.856103
CNVR_0395	0.88703783	**CNVR_2143**	0.768073	**CNVR_1027**	0.858799
CNVR_1383	0.88736818	**CNVR_2213**	0.771685	**CNVR_0637**	0.863866
CNVR_1879	0.88862702	**CNVR_2232**	0.776606	**CNVR_0164**	0.864131

Supp. Tbl. 2.1. Continued

CNVR_0107	0.889018782	**CNVR_2107**	0.777316	**CNVR_0531**	0.869072
CNVR_0223	0.890253115	**CNVR_1647**	0.7803	**CNVR_0894**	0.872476
CNVR_2162	0.890616769	**CNVR_1755**	0.780401	**CNVR_1188**	0.877086
CNVR_1582	0.89197037	**CNVR_0875**	0.782646	**CNVR_1755**	0.879525
CNVR_1058	0.895158089	**CNVR_0219**	0.786218	**CNVR_1582**	0.88025
CNVR_1277	0.898560431	**CNVR_1971**	0.786325	**CNVR_0905**	0.882863
CNVR_1121	0.901750773	**CNVR_2162**	0.792207	**CNVR_0245**	0.887725
CNVR_1974	0.902722661	**CNVR_0149**	0.795654	**CNVR_1458**	0.888601
CNVR_1164	0.904176134	**CNVR_0531**	0.79588	**CNVR_1343**	0.890231
CNVR_1577	0.905743642	**CNVR_2014**	0.797264	**CNVR_1030**	0.891142
CNVR_1707	0.912748429	**CNVR_0269**	0.798881	**CNVR_0010**	0.891325
CNVR_1533	0.914088698	**CNVR_1568**	0.806497	**CNVR_1295**	0.894691
CNVR_0616	0.91507572	**CNVR_1318**	0.807135	**CNVR_0558**	0.89571
CNVR_1130	0.915570464	**CNVR_1203**	0.807294	**CNVR_0548**	0.89627
CNVR_1343	0.91889468	**CNVR_2245**	0.815749	**CNVR_1047**	0.896635
CNVR_1624	0.92098975	**CNVR_1656**	0.819317	**CNVR_0094**	0.897196
CNVR_1592	0.922903057	**CNVR_0911**	0.822117	**CNVR_0462**	0.898941
CNVR_1119	0.923487632	**CNVR_1198**	0.829083	**CNVR_1077**	0.899951
CNVR_1677	0.925367677	**CNVR_1013**	0.834603	**CNVR_0493**	0.900721
CNVR_1318	0.926214361	**CNVR_1891**	0.83474	**CNVR_1820**	0.903388
CNVR_2143	0.930035109	**CNVR_1313**	0.83723	**CNVR_0816**	0.906345
CNVR_2124	0.93267596	**CNVR_1662**	0.839253	**CNVR_1318**	0.906884
CNVR_1755	0.935653015	**CNVR_0548**	0.840829	**CNVR_1088**	0.910758
CNVR_0531	0.936740524	**CNVR_2145**	0.849247	**CNVR_1156**	0.911315
CNVR_0320	0.937689089	**CNVR_0145**	0.851465	**CNVR_1532**	0.913872
CNVR_0118	0.941545369	**CNVR_1272**	0.862941	**CNVR_0403**	0.914938
CNVR_0639	0.942104203	**CNVR_0403**	0.865392	**CNVR_0125**	0.914961
CNVR_0940	0.942352407	**CNVR_0217**	0.867917	**CNVR_1677**	0.917861
CNVR_0919	0.94323468	**CNVR_1592**	0.869862	**CNVR_0063**	0.919702
CNVR_0470	0.94462729	**CNVR_1130**	0.870269	**CNVR_0708**	0.921044
CNVR_1783	0.945148182	**CNVR_1670**	0.880986	**CNVR_2014**	0.927443
CNVR_1859	0.945514313	**CNVR_1636**	0.882907	**CNVR_1502**	0.928301
CNVR_1935	0.946367731	**CNVR_0075**	0.884186	**CNVR_0034**	0.931084
CNVR_0251	0.954313117	**CNVR_1395**	0.888799	**CNVR_1921**	0.931488
CNVR_1730	0.957559954	**CNVR_0322**	0.901778	**CNVR_0295**	0.93195
CNVR_1395	0.957874013	**CNVR_2024**	0.904088	**CNVR_0056**	0.933616
CNVR_2030	0.958773916	**CNVR_0216**	0.905012	**CNVR_1246**	0.937479
CNVR_2047	0.962693856	**CNVR_1307**	0.907577	**CNVR_1090**	0.93878
CNVR_0403	0.964459952	**CNVR_0689**	0.914138	**CNVR_0885**	0.941457

Supp. Tbl. 2.1. Continued

CNVR_1652	0.966142639	**CNVR_0019**	0.919754	**CNVR_0096**	0.943959
CNVR_1983	0.971724513	**CNVR_1108**	0.923862	**CNVR_1700**	0.945451
CNVR_1308	0.971962883	**CNVR_2047**	0.923913	**CNVR_1370**	0.947664
CNVR_0610	0.972868084	**CNVR_0761**	0.925139	**CNVR_0956**	0.948274
CNVR_0189	0.97309586	**CNVR_1187**	0.927993	**CNVR_1138**	0.949686
CNVR_0052	0.973380333	**CNVR_1513**	0.950928	**CNVR_0084**	0.949711
CNVR_0571	0.974735735	**CNVR_1268**	0.953117	**CNVR_1599**	0.953029
CNVR_1494	0.975649412	**CNVR_2200**	0.955177	**CNVR_1829**	0.958263
CNVR_0498	0.976654834	**CNVR_2103**	0.955489	**CNVR_0949**	0.961667
CNVR_0502	0.981937262	**CNVR_1677**	0.956386	**CNVR_1355**	0.962281
CNVR_1836	0.982250513	**CNVR_1821**	0.957628	**CNVR_1214**	0.962606
CNVR_0895	0.983803453	**CNVR_1555**	0.957847	**CNVR_1315**	0.965
CNVR_2202	0.984125236	**CNVR_0018**	0.959122	**CNVR_0512**	0.9664
CNVR_1914	0.986553435	**CNVR_0137**	0.961777	**CNVR_1233**	0.967656
CNVR_1614	0.989448322	**CNVR_1841**	0.968581	**CNVR_2162**	0.970014
CNVR_0785	0.991500657	**CNVR_0570**	0.971577	**CNVR_1148**	0.973925
CNVR_0224	0.992043783	**CNVR_0247**	0.971895	**CNVR_0128**	0.978455
CNVR_1656	0.994289114	**CNVR_0184**	0.976977	**CNVR_1990**	0.987382
CNVR_1444	0.994409653	**CNVR_2148**	0.985161	**CNVR_1900**	0.987913
CNVR_0261	0.995094473	**CNVR_1826**	0.99405	**CNVR_1788**	0.990494
CNVR_1167	0.998420039	**CNVR_1956**	0.996512	**CNVR_1185**	0.992993
CNVR_1854	0.999899578	**CNVR_0165**	0.998806	**CNVR_1988**	0.998016

APPENDIX 2.2

Supplementary Table 2.2. Joint analysis of *TRPM2* SNP UKUL3936 (Comparison 1). Results of joint analysis comparing clinical foals with the combined subclinical and unaffected foals (comparison 1).

Genotype				
Standard Model	**Clinical foals**	**Subclinical + Unaffected foals**	**P value**	**Odds Ratio (95% CI)**
AA	72% (31/43)	43% (89/205)	NA	1 (NA)
AB	23% (10/43)	46% (95/205)	0.0016	0.29 (0.11 to 0.62)
BB	5% (2/43)	11% (21/205)	0.0826	0.27 (0.06 to 1.18)
Dominant Model				
Not AA	28% (12/43)	57% (116/205)	NA	1 (NA)
AA	72% (31/43)	43% (89/205)	0.0007	3.50 (1.71 to 7.16)
Additive Model				
f(A)*	2(0 to 2)	1(0 to 2)	0.0019	2.71 (1.45 to 5.06)
	84% (72/86)	66% (271/410)		

*** Median (range) reported for frequency of allele A, along with the proportion of A alleles among all alleles represented for each group. Joint analysis includes genotypes derived from SNP array and PCR genotyping.**

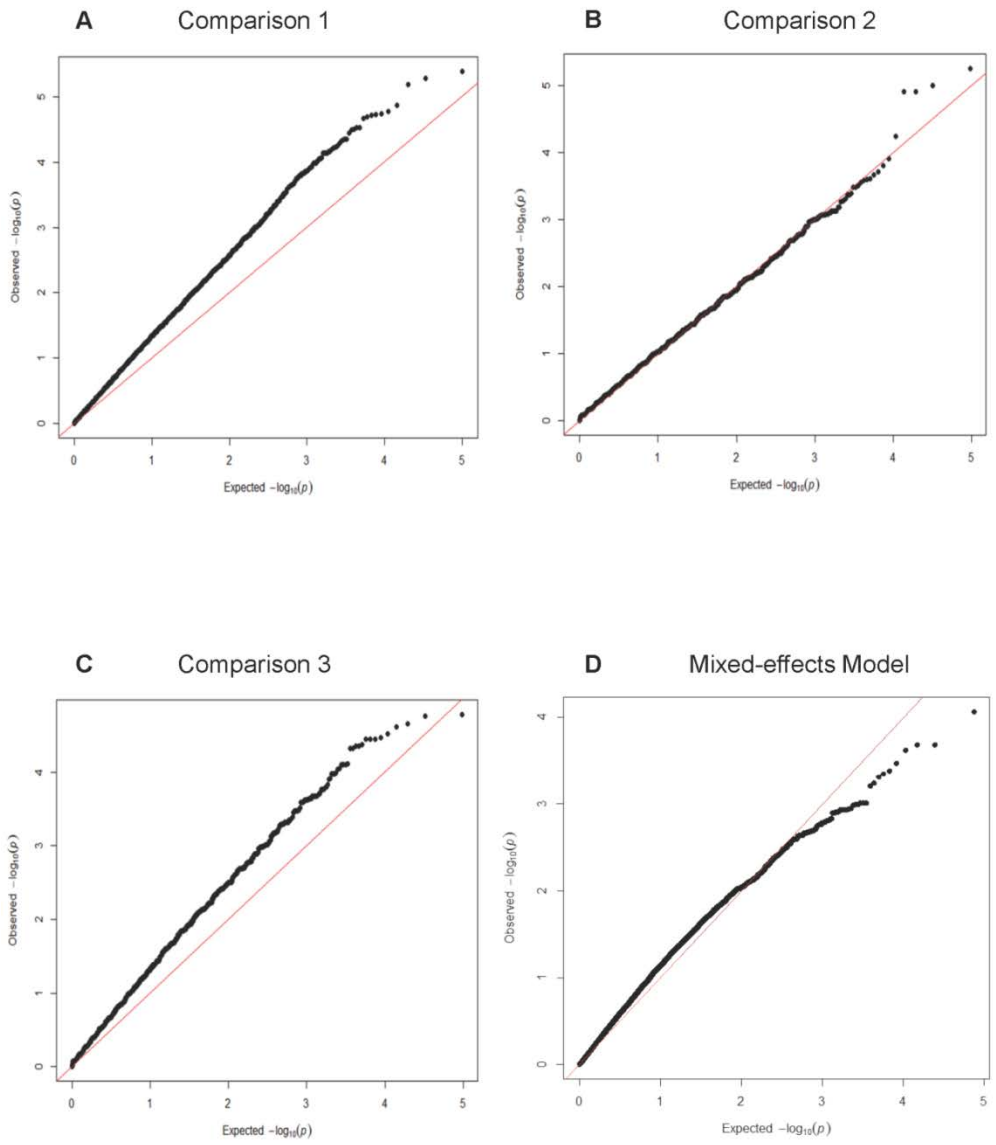
APPENDIX 2.3

Supplementary Table 2.3. Joint analysis of *TRPM2* SNP UKUL3936 (Comparison 3). Results of the joint analysis comparing clinical foals with unaffected foals.

Genotype				
Standard Model	**Clinical foals**	**Unaffected foals**	**P value**	**Odds Ratio (95% CI)**
AA	72% (31/43)	47% (23/49)	NA	1 (NA)
AB	23% (10/43)	47% (23/49)	0.0177	0.32 (0.13 to 0.81)
BB	5% (2/43)	6% (3/49)	0.4622	0.49 (0.08 to 3.20)
Dominant Model				
Not AA	28% (12/43)	47% (23/49)	NA	1 (NA)
AA	72% (31/43)	53% (26/49)	0.0179	2.92 (1.22 to 6.98)
Additive Model				
f(A)*	2(0 to 2)	1(0 to 2)	0.0398	2.20 (1.05 to 4.64)
	84% (72/86)	70% (69/98)		

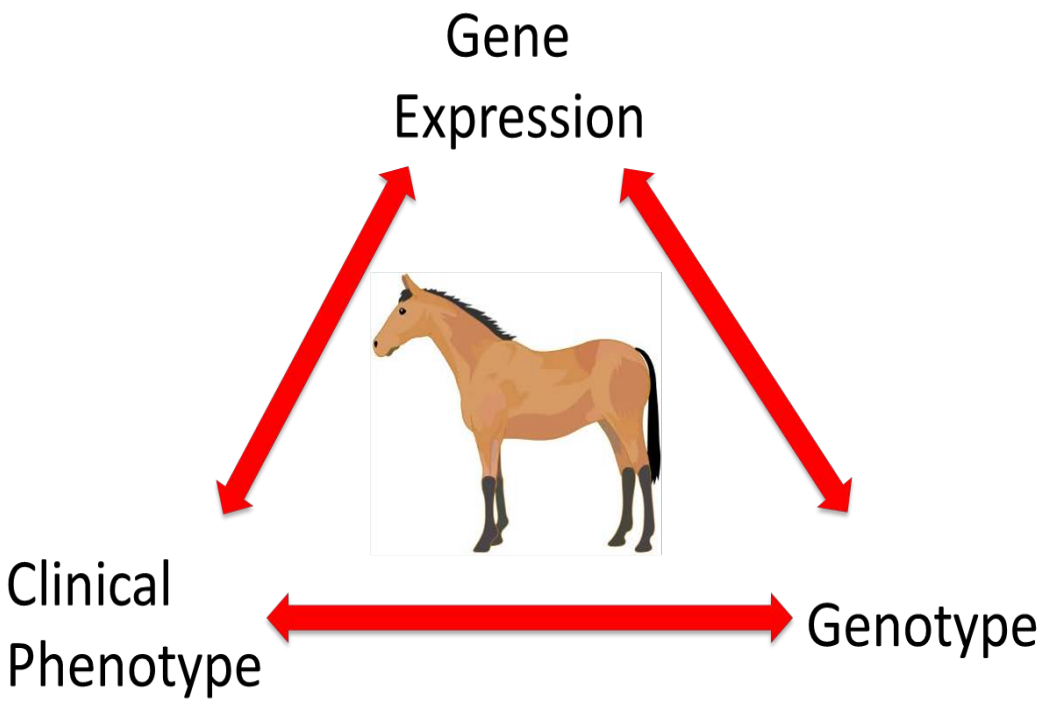
*** Median (range) reported for frequency of allele A, along with the proportion of A alleles among all alleles represented for each group. Joint analysis includes genotypes derived from SNP array and PCR genotyping.**

APPENDIX 2.4



Supplementary Figure 2.1. Quantile-quantile (QQ) plots of expected chi-squared significance values plotted against the observed values. QQ plots for (A) comparison 1, (B) comparison 2, (C) comparison 3, and (D) mixed-effects model.

APPENDIX 3.1



Supplementary Figure 3.1. Gene expression, clinical phenotype, and genotype relationship.